CANADIAN JOURNAL OF MATHEMATICS

Journal Canadien de Mathématiques

VOL. VIII · NO. 3

1956

Published for

THE CANADIAN MATHEMATICAL CONGRESS

by the

University of Toronto Press

The chief languages of the *Journal* are English and French.

Manuscripts for publication in the *Journal* should be sent to the *Editor-in-Chief*, H. S. M. Coxeter, University of Toronto. Everything possible should be done to lighten the task of the reader; the notation and reference system should be carefully thought out. Every paper should contain an introduction summarizing the results as far as possible in such a way as to be understood by the non-expert.

All other correspondence should be addressed to the *Managing Editor*, G. de B. Robinson, University of Toronto.

The *Journal* is published quarterly. Subscriptions should be sent to the *Managing Editor*. The price per volume of four numbers is $8.00. This is reduced to $4.00 for individual members of recognized Mathematical Societies.

The Canadian Mathematical Congress gratefully acknowledges the assistance of the following towards the cost of publishing this Journal:

University of Alberta

Assumption College	University of British Columbia
Carleton College	Ecole Polytechnique
Université Laval	Loyola College
University of Manitoba	McGill University
McMaster University	Université de Montréal
Queen's University	Royal Military College
St. Mary's University	University of Toronto

National Research Council of Canada
and the
American Mathematical Society

AUTHORIZED AS SECOND CLASS MAIL, POST OFFICE DEPARTMENT, OTTAWA

e
g
n
r
e

g

t
s
f

e
:

NOMBRES EXPONENTIELS ET NOMBRES DE BERNOULLI

JACQUES TOUCHARD

Introduction. Les nombres entiers positifs $a_0, a_1, \ldots, a_n, \ldots$ définis par la fonction génératrice

$$e^{e^z-1} = \sum_0^\infty a_n \frac{z^n}{n!},$$

et que l'on appelle nombres exponentiels jouent, pour la sommation de certaines séries, un rôle qui rappelle le rôle sommatoire des nombres de Bernoulli. Nous avons rassemblé ici les principales propriétés des nombres a_n dont plusieurs sont, croyons nous, nouvelles. L'une d'elles qui se présente spontanément, comme on le verra, est l'existence d'une orthogonalité symbolique par rapport à ces nombres. C'est ce qui nous a conduit à rechercher si l'on pouvait former une suite de polynômes orthogonaux symboliquement par rapport aux nombres de Bernoulli. On y parvient en effet, mais beaucoup moins facilement, grâce à une fraction continue donnée par Stieltjes.

Nous étudions aussi des polynômes exponentiels.

Nous ferons constamment usage du calcul symbolique, appelé calcul de Blissard. La plupart des démonstrations sont si simples qu'il suffira le plus souvent de les esquisser.

Propriétés des nombres exponentiels

1. En différentiant la formule symbolique

$$e^{e^z-1} = e^{az}, \tag{1}$$

on obtient

$$a_{n+1} = (a+1)^n \tag{2}$$

et, en posant dans (1), $e^z - 1 = u$, on trouve

$$a(a-1)\ldots(a-n+1) = 1. \tag{3}$$

Plus généralement, $f(u)$ désignant un polynôme quelconque,

$$\begin{aligned} f(a+1) &= a f(a), \\ f(a+p) &= a(a-1)\ldots(a-p+1) f(a). \end{aligned} \tag{4}$$

L'expression de $f(u)$ par la formule de Newton donne, en vertu de (3)

$$f(a) = f(0) + \frac{\Delta f(0)}{1} + \frac{\Delta^2 f(0)}{2!} + \ldots \tag{5}$$

et aussi

$$f(a+x) = f(x) + \frac{\Delta f(x)}{1} + \frac{\Delta^2 f(x)}{2!} + \ldots,$$

les différences étant prises pour la suite des valeurs x, $x+1$, $x+2, \ldots$ de la variable. En particulier

$$(6) \qquad \begin{aligned} a_p &= \frac{\Delta 0^p}{1} + \frac{\Delta^2 0^p}{2!} + \ldots + \frac{\Delta^p 0^p}{p!}, & p \geqslant 1,\\ a_{p+1} &= 1^p + \frac{\Delta 1^p}{1} + \frac{\Delta^2 1^p}{2!} + \ldots + \frac{\Delta^p 1^p}{p!}, & p \geqslant 0. \end{aligned}$$

Dans (5), substituons les expressions des différences successives, savoir

$$\Delta^p f(0) = f(p) - \binom{p}{1} f(p-1) + \binom{p}{2} f(p-2) - \ldots,$$

admettons ensuite que le polynôme $f(u)$ soit de degré $\leqslant n$ et posons

$$w(n) = 1 - \frac{1}{1} + \frac{1}{2!} - \frac{1}{3!} + \ldots + \frac{(-1)^n}{n!};$$

nous obtiendrons

$$f(a) = w(n) f(0) + \ldots + w(n-i) \frac{f(i)}{i!} + \ldots + w(0) \frac{f(n)}{n!}$$

et, en faisant grandir n indéfiniment,

$$e f(a) = \sum_{n=0}^{\infty} \frac{f(n)}{n!}.$$

On voit d'ailleurs directement, en multipliant les deux membres de (1) par e^{xz} et en développant que

$$e(a+x)^n = \sum_{i=0}^{\infty} \frac{(x+i)^n}{i!}$$

et que, $f(u)$ désignant un polynôme quelconque,

$$(7) \qquad e f(a+x) = \sum_{n=0}^{\infty} \frac{f(x+n)}{n!}.$$

C'est là la propriété sommatoire des nombres exponentiels. En particulier

$$(8) \qquad \begin{cases} e\, a_{n+1} = 1^n + \dfrac{2^n}{1} + \dfrac{3^n}{2!} + \ldots, & n \geqslant 0,\\ e\, a_0 = \displaystyle\sum_0^{\infty} \frac{1}{n!} \end{cases}$$

Les valeurs suivantes sont empruntées à Bell **(4)**. Le calcul a été poursuivi par Becker **(1)** jusqu'à l'indice $n = 35$ et par Miksa **(6a**, p. 54**)** jusqu'à l'indice $n = 51$.

n	a_n	n	a_n
0	1		
1	1	6	203
2	2	7	877
3	5	8	4140
4	15	9	21147
5	52	10	115975

2. Remplaçons maintenant, dans l'équation (5), $f(x)$ par $\Delta^n f(x)$ et faisons usage, au premier membre, de la formule (4) pour $p = 0, 1, 2, \ldots, n$, nous aurons symboliquement

$$h_n(a)\, f(a) = \Delta^n f(0) + \frac{\Delta^{n+1} f(0)}{1!} + \frac{\Delta^{n+2} f(0)}{2!} + \ldots,$$

où

$$h_n(x) = \sum_{i=0}^{n} (-1)^i \binom{n}{i} x(x-1)\ldots(x-n+i+1). \tag{9}$$

On en déduit que

$$a^p h_n(a) = \begin{cases} 0, & 0 < p < n, \\ n!, & p = n. \end{cases} \tag{10}$$

Grâce à ces formules on peut, connaissant $a_0, a_1, a_2, \ldots, a_{n-1}$ et le seul polynôme $h_n(x)$, calculer $a_n, a_{n+1}, \ldots, a_{2n}$. On voit de plus que les polynômes $h_n(x)$ jouissent de la belle propriété d'être orthogonaux symboliquement par rapport aux nombres a_n:

$$h_m(a)\, h_n(a) = \begin{cases} 0, & m \neq n, \\ m!, & m = n. \end{cases} \tag{11}$$

D'après (9), la fonction génératrice des polynômes $h_n(x)$ est

$$(1+z)^x e^{-z} = \sum_{n=0}^{\infty} h_n(x) \frac{z^n}{n!}.$$

En faisant, dans (7), $x = 0$ et $f(u) = h_m(u) h_n(u)$, on voit, d'après (11), que

$$\sum_{k=0}^{\infty} \frac{h_m(k)\, h_n(k)}{k!} = \begin{cases} 0, & m \neq n, \\ em!, & m = n. \end{cases}$$

C'est là un résultat qui se rattache à la théorie des polynômes orthogonaux de Charlier-Poisson **(10)**:

$$\begin{aligned}
h_0 &= 1, \\
h_1 &= x - 1, \\
h_2 &= x^2 - 3x + 1, \\
h_3 &= x^3 - 6x^2 + 8x - 1, \\
h_4 &= x^4 - 10x^3 + 29x^2 - 24x + 1, \\
h_5 &= x^5 - 15x^4 + 75x^3 - 145x^2 + 89x - 1, \\
h_6 &= x^6 - 21x^5 + 160x^4 - 545x^3 + 814x^2 - 415x + 1.
\end{aligned}$$

3. Les propriétés exposées dans les deux paragraphes précédents sont fondamentales. En voici d'autres.

Dans (1), posons $e^z = 1/(1-x)$ et divisons par $(1-x)$, nous aurons

$$(1-x)^{-1}e^{x/(1-x)} = (1-x)^{-a-1}.$$

Le premier membre se développe au moyen des polynômes de Laguerre, d'où

$$(a+1)(a+2)\dots(a+n) = \sum_{i=0}^{n}\binom{n}{i}^2 i!. \tag{12}$$

Cette formule résulte aussi de la relation

$$(a+x)(a+x-1)\dots(a+x-n+1) = \sum_{i=0}^{n}\binom{n}{i}x(x-1)\dots(x-n+1).$$

Dans (1), posons $e^z = (1+x)^2$, nous aurons

$$e^{2x+x^2} = (1+x)^{2a}.$$

Le premier membre se développe au moyen des polynômes d'Hermite, d'où

$$2a(2a-1)\dots(2a-n+1) = n!\sum_i \frac{1}{i!}\frac{2^{n-2i}}{(n-2i)!},$$

la somme s'étendant aux valeurs de i depuis zéro jusqu'à l'entier de $\frac{1}{2}n$.

Faisons maintenant, dans (1), $e^z = (1-x)^{\frac{1}{2}}$, il vient

$$e^{(1-x)^{\frac{1}{2}}-1} = (1-x)^{\frac{1}{2}a}. \tag{13}$$

Si l'on pose, pour un moment,

$$y(x) = -e^{(1-x)^{\frac{1}{2}}-1} = \sum_0^\infty \frac{c_n}{2^n}\frac{x^n}{n!}, \tag{14}$$

la fonction $y(x)$ satisfait à l'équation différentielle

$$(4-4x)\,y''(x) = 2\,y'(x) + y(x),$$

d'où se tire la récurrence

$$\begin{array}{ll} c_{n+2} = (2n+1)\,c_{n+1} + c_n, & \\ c_0 = -1, & c_6 = 329, \\ c_1 = 1, & c_7 = 3655, \\ c_2 = 0, & c_8 = 47844, \\ c_3 = 1, & c_9 = 721315, \\ c_4 = 5, & c_{10} = 12310199, \\ c_5 = 36, & c_{11} = 234615096, \end{array} \tag{15}$$

et l'on voit, d'après (15), qu'à partir de $n = 4$ les nombres c_n sont les dénominateurs des réduites de la fraction continue

$$\frac{1|}{|5} + \frac{1|}{|7} + \frac{1|}{|9} + \frac{1|}{|11} + \dots$$

qui est une fraction continue de Gauss. On en déduit l'expression de c_{n+4}, pour $n > 0$

$$c_{n+4} = 5\cdot 7\cdot 9 \ldots (2n+5)\left[1 + \binom{n}{1}\frac{1}{5\cdot(2n+5)}\right.$$

$$(16) \qquad + \binom{n-1}{2}\frac{1}{5\cdot 7\cdot(2n+3)\cdot(2n+5)}$$

$$\left. + \binom{n-2}{3}\frac{1}{5\cdot 7\cdot 9\cdot(2n+1)(2n+3)(2n+5)} + \ldots\right].$$

Les formules (13) et (14) donnent alors

$$(17) \qquad a(a-2)(a-4)\ldots(a-2k+2) = (-1)^{k-1}c_k.$$

Par analogie avec la formule (3), considérons l'intégrale

$$\frac{(n-1)!}{(a+1)(a+2)\ldots(a+n)} = \int_0^1 (1-x)^a x^{n-1} dx.$$

Comme on a symboliquement

$$(1-x)^a = e^{-x},$$

cette égalité devient

$$(18) \qquad \frac{(n-1)!}{(a+1)(a+2)\ldots(a+n)} = \int_0^1 e^{-x} x^{n-1} dx = P(n),$$

$P(x)$ désignant la fonction bien connue de Prÿm. On a donc

$$\frac{1}{(a+1)(a+2)\ldots(a+n)} = \frac{P(n)}{\Gamma(n)} = e^{-1}\left[e - 1 - \frac{1}{1} - \frac{1}{2!} - \ldots - \frac{1}{(n-1)!}\right]$$

ou bien, d'après une propriété connue de $P(z)$,

$$\frac{e}{(a+1)(a+2)\ldots(a+n)} = \sum_{i=1}^{\infty} \frac{1}{\Gamma(n+i)},$$

ce qui s'accorde avec la formule (7), quand on y fait $x = 0$ et $f(u) = \Gamma(u+1)/\Gamma(u+n+1)$. Il est facile de démontrer que si l'on développe le premier membre de (18) suivant les puissances positives de a et si l'on somme, par la méthode de Borel, la série divergente obtenue en remplaçant a^n par a_n, on obtient $P(n)$.

Considérons enfin les nombres

$$q_n = a_n a_0 - \binom{n}{1} a_{n-1}a_1 + \binom{n}{2} a_{n-2}a_2 - \ldots + (-1)^n a_0 a_n$$

ou, symboliquement,

$$(19) \qquad q_n = (a - a')^n.$$

Ces nombres q_n sont les invariants quadratiques des formes binaires $(x + ay)^n$. Ceux d'indice impair sont nuls. D'après (19), leur fonction génératrice est

$$e^{e^z+e^{-z}-2} = \sum_0^{\infty} q_n \frac{z^n}{n!} = e^{qz} \tag{20}$$

et, en différentiant cette équation, on obtient la récurrence très simple

$$q_{n+1} = (q+1)^n - (q-1)^n,$$

qui donne

$$\begin{array}{ll} q_0 = 1, & q_8 = 3614, \\ q_2 = 2, & q_{10} = 99302, \\ q_4 = 14, & q_{12} = 3554894, \\ q_6 = 182, & q_{14} = 159175382. \end{array}$$

Dans (20), changeons z en $2\,iz$, remarquons que

$$e^{2iz} + e^{-2iz} - 2 = -4\sin^2 z$$

et posons $\sin z = \frac{1}{2}u$, nous obtenons, puisque $q_{2n+1} = 0$, l'égalité symbolique

$$e^{-u^2} = \cos\left(2\,q \arcsin\frac{u}{2}\right). \tag{21}$$

Or on sait que

$$\cos\left(2\,m \arcsin\frac{u}{2}\right) = 1 - \frac{m^2}{2!}u^2 + \frac{m^2(m^2-1^2)}{4!}u^4 - \dots.$$

Le développement des deux membres de (21) conduit donc à la formule

$$q^2(q^2-1^2)(q^2-2^2)\dots[q^2-(k-1)^2] = \frac{(2k)!}{k!}. \tag{22}$$

Les nombres q_n jouissent d'une propriété sommatoire que nous établirons plus loin.

4. Une définition intéressante des nombres exponentiels a_n a été donnée par Broggi **(5)** au moyen de la série asymptotique

$$I(x) = e^{-1}\int_0^1 t^{x-1}e^t dt = \sum_{i=0}^{\infty}\frac{(-1)^i a_i}{x^{i+1}}, \qquad R(x) > 0 \tag{23}$$

qu'on obtient en développant d'abord e^t, en intégrant terme à terme, en développant ensuite les fractions suivant les puissances de $1/x$ et en ayant égard aux formules (8). Or la fonction $\mathrm{I}(x)$ peut être représentée par la fraction continue

$$I(x) = \frac{1|}{|x+1} - \frac{1|}{|x+2} - \frac{2|}{|x+3} - \frac{3|}{|x+4} - \dots$$

En désignant les réduites par α_n/β_n, on a

$$\frac{\alpha_0}{\beta_0} = \frac{0}{1},\quad \frac{\alpha_1}{\beta_1} = \frac{1}{x+1}, \dots$$

et, en général, α_n est un polynôme de degré $n-1$, β_n un polynôme de degré n, qui satisfont tous les deux à l'équation aux différences

$$u_n = (x+n)\,u_{n-1} - (n-1)\,u_{n-2}.$$

Partant de cette relation et des valeurs initiales $\beta_0 = 1$, $\beta_1 = x+1$, un calcul facile montre que

$$\sum_0^\infty \beta_n(x)\frac{z^n}{n!} = (1-z)^{-x} e^z,$$

d'où se tire l'expression

$$\beta_n(x) = (-1)^n h_n(-x),$$

$h_n(x)$ étant le polynôme (9). Il résulte alors de la théorie des fractions continues que

$$\frac{\alpha_n(x)}{\beta_n(x)} = \frac{a_0}{x} - \frac{a_1}{x^2} + \ldots - \frac{a_{2n-1}}{x^{2n}} + \frac{1}{x^{2n+1}} R_n\left(\frac{1}{x}\right),$$

$R_n(1/x)$ désignant une série en $1/x$, et les propriétés d'orthogonalité symbolique des polynômes $h_n(x)$, que nous avons rencontrées au §2, sont, comme il est aisé de le voir, une conséquence immédiate de cette formule. Nous utiliserons cette remarque plus loin.

5. Epstein **(5)** a considéré la fonction entière de s

$$g(s) = \frac{1}{1^s} + \frac{1}{1}\frac{1}{2^s} + \frac{1}{2!}\frac{1}{3^s} + \ldots. \tag{24}$$

On a évidemment

$$\begin{cases} g(-n) = e\,a_{n+1}, \\ g(1) \;\;= e-1 \end{cases} \qquad n > 0, \tag{25}$$

et, d'après une intégrale eulérienne classique,

$$g(n) = \frac{(-1)^{n-1}}{(n-1)!}\int_0^1 e^t \log^{n-1} t\, dt, \qquad n > 1 \tag{26}$$

formule que nous généraliserons plus loin. Epstein a calculé les valeurs de $e^{-1}g(n)$, pour n entier positif jusqu'à $n = 21$, mais, sauf pour $n = 0$ et $n = 1$, ces valeurs ne paraissent pas pouvoir s'exprimer à l'aide de nombres connus.

En posant $g(n) = g_n$, on vérifie sans peine la relation symbolique

$$(1-g)(1-2g)\ldots(1-ng)$$
$$= \frac{1}{(n+1)^n} + \frac{1}{1}\frac{1}{(n+2)^n} + \frac{1}{2!}\frac{1}{(n+3)^n} + \ldots$$

qui présente une certaine analogie avec (3) et constitue une formule de récurrence approchée des nombres g_n.

Polynômes exponentiels

6. Les polynômes $\phi_n(x)$ dont nous désirons nous occuper et qui se présentent souvent en analyse sont définis par la fonction génératrice

$$e^{x(e^z-1)} = \sum_{n=0}^{\infty} \phi_n(x)\,\frac{z^n}{n!}. \tag{27}$$

$$\begin{aligned}
\phi_0 &= 1,\\
\phi_1 &= x,\\
\phi_2 &= x^2 + x,\\
\phi_3 &= x^3 + 3\,x^2 + x,\\
\phi_4 &= x^4 + 6\,x^3 + 7\,x^2 + x,\\
\phi_5 &= x^5 + 10\,x^4 + 25\,x^3 + 15\,x^2 + x.
\end{aligned}$$

On peut aussi les définir par le développement asymptotique

$$e^{-x}\int_0^1 e^{xt}t^{z-1}dt = \frac{1}{z}\,\phi_0(x) - \frac{1}{z^2}\,\phi_1(x) + \ldots + \frac{(-1)^n}{z^{n+1}}\,\phi_n(x) + \ldots.$$

La plupart des propriétés des nombres a_n peuvent être étendues aux polynômes ϕ_n. Enumérons les principales:

$$\begin{aligned}
&\phi_n(1) = a_n,\\
&\phi_{n+1}(x) = x(\phi + 1)^n,\\
&\phi_{n+1}(x) = x(\phi_n + \phi_n'),\\
&\phi_n' = (\phi + 1)^n - \phi_n,\\
&\phi(\phi - 1)\ldots(\phi - n + 1) = x^n.
\end{aligned} \tag{28}$$

Si on pose $j_n(x) = x(x-1)\ldots(x-n+1)$, cette formule s'écrit symboliquement $j_n(\phi) = x^n$ et l'on a aussi $\phi_n(j) = x^n$. Les polynômes j_n *et* ϕ_n sont donc inverses les uns des autres.

Nous avons ensuite

$$\begin{aligned}
\phi_n(x) &= 0^n + \frac{x}{1}\,\Delta 0^n + \frac{x^2}{2!}\,\Delta^2 0^n + \ldots + \frac{x^n}{n!}\,\Delta^n 0^n,\\
e^{x}\phi_0(x) &= \sum_{s=0}^{\infty}\frac{x^s}{s!},\\
e^{x}\phi_{n+1}(x) &= x\sum_{s=0}^{\infty}(s+1)^n\frac{x^s}{s!}, \qquad n > 0.
\end{aligned}$$

Plus généralement, $f(u)$ désignant un polynôme quelconque de degré n, on a

$$\begin{aligned}
&x f(\phi + 1) = f(\phi),\\
&x^p f(\phi + p) = \phi(\phi - 1)\ldots(\phi - p + 1)\,f(\phi),
\end{aligned} \tag{29}$$

$$f(\phi) = f(0) + \frac{x}{1}\,\Delta f(0) + \frac{x^2}{2!}\,\Delta^2 f(0) + \ldots + \frac{x^n}{n!}\,\Delta^n f(0),$$

$$e^x f(\phi + k) = \sum_{s=0}^{\infty}\frac{x^s f(s+k)}{s!}. \tag{30}$$

Cette dernière formule, où k est un nombre quelconque, exprime la propriété sommatoire des polynômes ϕ_n. L'équation (28) est équivalente à

$$\frac{\phi_n(t)\, e^t}{t} = \frac{d}{dt}[\phi_{n-1}(t)\, e^t].$$

En utilisant cette relation et en intégrant plusieurs fois par parties le second membre de (26), on obtient, k étant un entier positif ou nul,

$$g(n) = \frac{(-1)^{n+k}}{(n+k)!}\int_0^1 \frac{\phi_{k+2}(t)}{t}\, e^t \log^{n+k} t\, dt, \qquad n > -k,$$

qui est un prolongement de l'intégrale (26) pour des valeurs négatives de n.

7. En raisonnant comme au §2 et en s'appuyant sur la formule (29), on verra que les polynômes

$$H_n(x, z) = z(z-1)\ldots(z-n+1) - \binom{n}{1} x\, z(z-1)\ldots(z-n+2)$$
$$+ \binom{n}{2} x^2 z(z-1)\ldots(z-n+3) - \ldots + (-1)^n x^n$$

satisfont aux relations symboliques

$$\begin{aligned} H_n(x, \phi) &= 0, \\ \phi H_n(x, \phi) &= 0, \\ &\cdots\cdots \\ \phi^{n-1} H_n(x, \phi) &= 0, \\ \phi^n H_n(x, \phi) &= n!\, x^n \end{aligned}$$

et que, par conséquent, ces polynômes sont orthogonaux symboliquement par rapport aux polynômes $\phi_n(x)$. En faisant usage de la formule (30) pour $k = 0$ et $f(u) = H_m(x, u)\cdot H_n(x, u)$ on aura

$$\sum_{s=0}^{\infty} \frac{x^s H_m(x, s)\, H_n(x, s)}{s!} = \begin{cases} 0, & m \neq n, \\ m!\, x^m e^x, & m = n. \end{cases}$$

On observera que ce sont là de pures identités et que l'on peut y remplacer les diverses puissances de x par des nombres arbitraires.

Notons encore que, par analogie avec les nombres q_n du §3, on peut considérer les polynômes définis par l'égalité

$$\chi_n(x) = \phi_n\phi_0 - \binom{n}{1}\phi_{n-1}\phi_1 + \binom{n}{2}\phi_{n-2}\phi_2 - \ldots + (-1)^n \phi_0\phi_n.$$

On démontrera comme au §3 la belle relation symbolique

$$\chi^2(\chi^2 - 1^2)(\chi^2 - 2^2)\ldots[\chi^2 - (k-1)^2] = \frac{(2k)!}{k!} x^k.$$

Propriétés arithmétiques

8. Les nombres exponentiels satisfont à diverses congruences dont les plus simples se tirent de (2), (3), (4) et de la congruence identique de Lagrange

$$x(x-1)\dots(x-p+1) \equiv x^p - x \pmod{p} \tag{31}$$

où p désigne, comme dans tout le reste de ce paragraphe, un nombre premier. Voici quelques unes de ces congruences

$$a_p \equiv 2, \quad a_{p+1} \equiv 3, \quad a_{p^v+h} \equiv v\, a_h + a_{h+1} \pmod{p}, \tag{32}$$

et notamment

$$a_{p^v} \equiv v + 1 \pmod{p}$$

relation qui montre que la suite des restes (mod p) de $a_1, a_p, a_{p^2}, \dots$ admet la période p.

D'après les formules (4) et (12) on a symboliquement

$$[a(a-1)\dots(a-p+1)]^2 = (a+1)(a+2)\dots(a+p) \equiv 1 + p! \pmod{p^2}$$

or, d'après le théorème de Wilson

$$p! + p \equiv 0 \pmod{p^2}$$

donc

$$[a(a-1)\dots(a-p+1)]^2 \equiv 1 - p \pmod{p^2}. \tag{33}$$

D'autre part, en vertu de (31),

$$[x(x-1)\dots(x-p+1)]^2 - 2x(x-1)\dots(x-p+1)(x^p - x) + (x^p - x)^2 \equiv 0 \pmod{p^2}.$$

En remplaçant x^k par a_k et en transformant le second terme à l'aide de (4), on obtient

$$[a(a-1)\dots(a-p+1)]^2 - 2(a_p - a_1 - p) + a_{2p} - 2a_{p+1} + a_2 \equiv 0 \pmod{p^2},$$

et en comparant à (33) on trouve la relation

$$a_{2p} - 2\, a_{p+1} - 2\, a_p + p + 5 \equiv 0 \pmod{p^2}.$$

Une autre congruence concerne les nombres c_k du §3 et se tire de la formule (17). On a évidemment, lorsque p est un nombre premier impair,

$$x(x-2)(x-4)\dots(x-2p+2) \equiv x^p - x \pmod{p}$$

et, par conséquent, d'après (17) et (32)

$$c_p \equiv 1 \pmod{p}, \qquad p \text{ premier impair.}$$

On a ensuite, d'après (15),

$$c_{p+2} - c_{p+1} \equiv 1 \pmod{p}.$$

Considérons maintenant les nombres q_n du §3 et supposons encore p premier impair. On sait que

$$x(x-1^2)(x-2^2)\dots\left[x - \left(\frac{p-1}{2}\right)^2\right] \equiv x^{\frac{1}{2}(p+1)} - x \pmod{p}. \tag{34}$$

D'autre part, dans la formule (22), faisons $k = \frac{1}{2}(p + 1)$; le second membre devient divisible par p. Donc, en faisant dans (34) $x = q^2$ symboliquement, on obtient

$$q_{p+1} - q_2 \equiv 0 \qquad (\text{mod } p),$$

ou bien

(35) $$q_{p+1} \equiv 2 \pmod p, \qquad p \text{ premier impair.}$$

Maintenant, on a aussi identiquement

$$(x - 1^2)(x - 2^2) \ldots [x - (p - 1)^2] \equiv (x^{\frac{1}{2}(p-1)} - 1)^2 \qquad (\text{mod } p);$$

d'où

(36) $$x(x - 1^2) \ldots [x - (p - 1)^2] \equiv x^p - 2x^{\frac{1}{2}(p+1)} + x \qquad (\text{mod } p);$$

mais si dans (22) on fait $k = p$ le second membre est divisible par p. On a donc, en remplaçant dans (36) x par q^2 symboliquement

$$q_{2p} - 2\, q_{p+1} + q_2 \equiv 0 \qquad (\text{mod } p),$$

et, en vertu de (35)

(37) $$q_{2p} \equiv 2 \pmod p, \qquad p \text{ premier impair.}$$

En comparant (35) et (37) on voit que, si p et $2p - 1$ sont tous deux premiers impairs, on a

$$q_{2p} \equiv 2 \qquad (\text{mod } p(2p - 1)),$$

congruence que l'on peut vérifier pour $p = 3$ et $p = 7$ et qui a lieu aussi pour $p = 2$.

Les polynômes $\phi_n(x)$ satisfont aussi à diverses congruences, soit qu'on considère x comme un nombre entier, soit plutôt qu'on considère x comme une indéterminée, racine d'une congruence irréductible (mod p), c'est-à-dire comme une imaginaire de Galois. Nous renverrons à ce sujet à **(12)**.

Intégrales définies

9. On peut de plusieurs manières obtenir l'expression des nombres a_n, des polynômes ϕ_n et de la fonction $g(s)$ par des intégrales définies. En voici quelques unes.

L'intégrale eulérienne de deuxième espèce donne d'abord

$$\Gamma(s)\, g(s) = \int_0^\infty e^{-x} e^{e^{-x}} x^{s-1} dx, \qquad R(s) > 0$$

et on en tire par un procédé connu la formule, valable dans tout le plan s,

$$\frac{g(1 - s)}{\Gamma(s)} = \frac{1}{2\pi i} \int_C e^z e^{e^z} z^{-s} dz$$

où C désigne un lacet partant de $-\infty$ avec l'argument $-\pi$ pour z et y revenant avec l'argument $+\pi$, après avoir entouré l'origine.

En partant d'une formule, dûe à Laplace,

$$\int_{-\infty}^{+\infty} \frac{e^{iuz}dz}{(1+iz)^k} = \begin{cases} \dfrac{2\pi}{\Gamma(k)} u^{k-1}e^{-u}, & u > 0, \\ 0, & u < 0, \end{cases}$$

et où k est positif, on obtient à l'aide de (8), après quelques calculs

$$\frac{\pi a_n}{\Gamma(n)} = \int_0^\infty \frac{e^{\cos z}}{(1+z^2)^{\frac{1}{2}n}} \cos[z + e \sin z - n \operatorname{arctg} z]\, dz, \qquad n > 1.$$

10. Si l'on considère la distribution de Poisson

$$d\alpha(x, t) = \frac{e^{-x}x^t}{t!}, \qquad t = 0, 1, 2, 3, \ldots \tag{38}$$

c'est-à-dire la "step-function" ayant le saut $d\,\alpha(x, t)$ aux points $t = 0, 1, 2, 3, \ldots$, on obtient l'intégrale de Stieltjes

$$e^{x(e^z-1)} = \int_0^\infty e^{zt} d\alpha(x, t)$$

et, par suite,

$$\phi_n(x) = \int_0^\infty t^n d\alpha(x, t), \qquad n > 0.$$

Les polynômes $\phi_n(x)$ sont donc les moments de la distribution (38). On a aussi, pour toute valeur de s,

$$\int_1^\infty t^{1-s} d\alpha(1, t) = e^{-1}g(s).$$

11. Mais l'expression la plus intéressante des nombres a_n paraît être celle qui se déduit de la formule **(11)**

$$e^x = \int_0^\infty \frac{x^z dz}{\Gamma(1+z)} + \int_0^\infty \frac{e^{-xz}dz}{z(\pi^2 + \log^2 z)}, \qquad R(x) > 0. \tag{39}$$

En remplaçant x par e^x, en différentiant n fois et en tenant compte de (6), on obtient

$$e\, a_n = \int_0^\infty \frac{z^n dz}{\Gamma(1+z)} - \frac{1}{\pi} \int_0^\infty \frac{e^{-z}\Gamma(n) \sin\{n \operatorname{arctg}(\pi/\log z)\}\, dz}{(\pi^2 + \log^2 z)^{\frac{1}{2}n}},$$

équation qui est exacte même pour $n = 0$.

Dans la formule (39), remplaçons x par $x\,e^t$ et développons suivant les puissances de t, nous obtiendrons

$$e^x\phi_n(x) = \int_0^\infty \frac{x^z z^n dz}{\Gamma(1+z)} + \int_0^\infty \frac{e^{-u}\phi_n(-u)\, du}{u\{\pi^2 + \log^2(x/u)\}}, \qquad R(x) > 0, n > 0.$$

Faisons dans cette équation, successivement $n = 0, 1, 2, \ldots, p$, rappelons nous que la factorielle $j_p(x)$ du §6 satisfait à la relation synbolique $j_p(\phi) = x^p$ et nous trouverons

$$e^x x^p = \int_{-p}^\infty \frac{x^{z+p}dz}{\Gamma(1+z)} + (-1)^p \int_0^\infty \frac{e^{-u}u^p du}{u\{\pi^2 + \log^2(x/u)\}}.$$

NOMBRES EXPONENTIELS À PLUSIEURS ARGUMENTS

12. Nous indiquerons rapidement une généralisation des nombres exponentiels a_n. Si on pose

$$a_n(\omega, \omega') = (\omega a + \omega' a')^n = \sum_{i=0}^{n} \binom{n}{i} \omega^{n-i} \omega'^i a_{n-i} a_i,$$

ω et ω' étant deux constantes, la fonction génératrice des nombres $a_n(\omega, \omega')$ sera

$$e^{e^{\omega z}-1} e^{e^{\omega' z}-1} = \sum_{n=0}^{\infty} a_n(\omega, \omega') \frac{z^n}{n!}$$

et, en multipliant les deux membres de cette égalité par e^{xz} et développant suivant les puissances de z, on verra que

$$\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \frac{f(x + m\omega + n\omega')}{m!\, n!} = e^2 f[x + a(\omega, \omega')] \tag{40}$$

où $f(u)$ désigne un polynôme arbitraire. On peut ainsi former des nombres a_n dépendant d'un nombre quelconque d'arguments $\omega, \omega', \omega'', \ldots$. En particulier, pour $\omega = 1$, $\omega' = -1$, on aura

$$a_k(1, -1) = q_k,$$

q_k désignant les nombres définis par (19) au §3 et, par suite

$$\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \frac{f(x + m - n)}{m!\, n!} = e^2 f(x + q)$$

symboliquement. C'est la propriété sommatoire des nombres q_n. On a notamment

$$e^2 q_{2k} = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \frac{(m-n)^{2k}}{m!\, n!}.$$

On peut généraliser d'une manière analogue les polynômes $\phi_n(x)$. En particulier, on verra que

$$\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \frac{x^{m+n}(m-n)^k}{m!\, n!} = e^{2x} \chi_k(x)$$

où $\chi_n(x)$ est le polynôme introduit au §7.

NOMBRES DE BERNOULLI

13. Les nombres de Bernoulli $b_0, b_1, b_2, \ldots$ sont définis par

$$\frac{x}{e^x - 1} = \sum_{0}^{\infty} \frac{b_n x^n}{n!}, \quad b_0 = 1, \quad b_1 = -\tfrac{1}{2}, \quad b_2 = \tfrac{1}{6}, \ldots ; \; b_{2n+1} = 0, \quad n > 1,$$

et nous cherchons une suite de polynômes $Q_n(x)$ orthogonaux symboliquement par rapport à ces nombres, c'est-à-dire tels que l'on ait symboliquement

$$b^p Q_n(b) = \begin{cases} 0, & 0 < p < n, \\ K_n, & p = n, \end{cases} \tag{41}$$

K_n étant une constante.

Il est nécessaire pour cela, comme nous l'avons remarqué au §4, de savoir mettre sous forme de fraction continue d'un type approprié (**7**, chap. 9) la série asymptotique

$$\psi(x+1) = \frac{b_0}{x} + \frac{b_1}{x^2} + \frac{b_2}{x^3} + \dots. \tag{42}$$

Les polynômes $Q_n(x)$ seront alors les dénominateurs des réduites successives. On sait par la théorie de la fonction Γ que

$$\psi(x+1) = \frac{d^2 \log \Gamma(x+1)}{dx^2} = \sum_{n=1}^{\infty} \frac{1}{(x+n)^2}, \tag{43}$$

et il se trouve que l'on déduit immédiatement d'un résultat de Stieltjes (**9**, p. 378) la fraction continue ayant le type qui convient

$$\psi(x+1) = \frac{2|}{|2x+1} + \frac{\lambda_1|}{|2x+1} + \frac{\lambda_2|}{|2x+1} + \frac{\lambda_3|}{|2x+1} + \dots,$$

où

$$\lambda_n = \frac{n^4}{4n^2-1}.$$

Les polynômes $Q_n(x)$ vérifient donc la formule de récurrence

$$Q_{n+1}(x) = (2x+1)\,Q_n(x) + \frac{n^4}{4n^2-1}\,Q_{n-1}(x) \tag{44}$$

à l'aide de laquelle nous avons calculé:

$$Q_0 = 1,$$

$$Q_1 = 2x+1,$$

$$Q_2 = 4\left(x^2 + x + \frac{1}{3}\right),$$

$$Q_3 = 4\left(2x^3 + 3x^2 + \frac{11}{5}x + \frac{3}{5}\right),$$

$$Q_4 = 16\left(x^4 + 2x^3 + \frac{17}{7}x^2 + \frac{10}{7}x + \frac{12}{35}\right),$$

$$Q_5 = 16\left(2x^5 + 5x^4 + \frac{80}{9}x^3 + \frac{25}{3}x^2 + \frac{274}{63}x + \frac{20}{21}\right),$$

$$Q_6 = 64\left(x^6 + 3x^5 + \frac{80}{11}x^4 + \frac{105}{11}x^3 + \frac{89}{11}x^2 + \frac{42}{11}x + \frac{60}{77}\right),$$

$$Q_7 = 64\left(2x^7 + 7x^6 + \frac{287}{13}x^5 + \frac{490}{13}x^4 + \frac{6559}{143}x^3 + \frac{4949}{143}x^2 + \frac{198}{13}x + \frac{420}{143}\right),$$

$$Q_8 = 256\left(x^8 + 4x^7 + \frac{238}{15}x^6 + \frac{168}{5}x^5 + \frac{2135}{39}x^4 + \frac{756}{13}x^3 \right.$$
$$\left. + \frac{88316}{2145}x^2 + \frac{12176}{715}x + \frac{448}{143}\right),$$

$$Q_9 = 256\left(2x^9 + 9x^8 + \frac{744}{17}x^7 + \frac{1890}{17}x^6 + \frac{19698}{85}x^5 + \frac{5481}{17}x^4 \right.$$
$$\left. + \frac{71756}{221}x^3 + \frac{47340}{221}x^2 + \frac{1026576}{12155}x + \frac{36288}{2431}\right).$$

Comme vérification de ces expressions, nous observerons que, d'après (44), on a

$$Q_{n+1}(-\tfrac{1}{2}) = \frac{n^4}{4n^2 - 1} Q_{n-1}(-\tfrac{1}{2})$$

d'où l'on déduit

$$\begin{aligned} &Q_{2n+1}(-\tfrac{1}{2}) = 0, \\ (45) \qquad &Q_{2n}(-\tfrac{1}{2}) = \frac{[1\cdot 3\cdot 5\cdots(2n-1)]^4}{1\cdot 3\cdot 5\cdot 7\cdots(4n-1)}. \end{aligned}$$

La constante K_n qui figure dans (41) s'obtient en multipliant les deux membres de (44) par x^{n-1} et en remplaçant x symboliquement par b, ce qui donne

$$0 = 2K_n + \frac{n^4}{4n^2 - 1} K_{n-1}$$

et comme $K_0 = b_0 Q_0(b) = 1$, on obtient

$$(46) \qquad K_n = \frac{(-1)^n}{2n+1} \frac{1}{2^n} \frac{[n!]^4}{[1\cdot 3\cdot 5\cdots(2n-1)]^2}.$$

Le coefficient de x^n dans $Q_n(x)$ étant égal à 2^n, on a les relations d'orthogonalité symbolique

$$(47) \qquad Q_m(b)\, Q_n(b) = \begin{cases} 0, & m \neq n, \\ 2^n K_n, & m = n, \end{cases}$$

où K_n a la valeur (46).

Si l'on considère la série

$$y(z) = \sum_{n=0}^{\infty} Q_n(x) \frac{z^n}{n!},$$

on peut former au moyen de (44), une équation différentielle linéaire du troisième ordre à laquelle elle satisfait et montrer qu'elle converge pour $|z| < 2$. C'est ce qu'on vérifie, en faisant $x = -\tfrac{1}{2}$, à l'aide de (45) et de la formule de Stirling. Mais l'expression générale des polynômes $Q_n(x)$ nous échappe.

14. Les formules symboliques (47) donnent naissance à une orthogonalité véritable de la manière suivante.

Il est facile de démontrer, au moyen du calcul des résidus et en vertu de (43), que

$$\psi(x+1) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \pi \frac{\cot \pi z}{(z-x)^2}\, dz, \qquad -1 < c < 0$$

le point x étant soit à droite soit à gauche de la droite d'intégration d'abscisse c. Intégrons par parties, nous obtenons

$$\psi(x+1) = \frac{\pi i}{2} \int_{c-i\infty}^{c+i\infty} \frac{1}{\sin^2 \pi z} \frac{dz}{z-x}.$$

En développant $1/(z-x)$ suivant les puissances de $1/x$, sans avoir égard à la convergence et en comparant à la série asymptotique (42), on est conduit à penser que

$$b_n = -\frac{\pi i}{2}\int_{c-i\infty}^{c+i\infty} \frac{z^n dz}{\sin^2 \pi z}, \qquad \begin{matrix} -1 < c < 0, \\ n = 0, 1, 2, \ldots . \end{matrix} \tag{48}$$

Or cette expression des nombres de Bernoulli qui parait être restée inaperçue ou, en tous cas, peu employée n'est en définitive, qu'une application d'une belle formule, établie par Jensen, pour la fonction $\zeta(s)$ de Riemann, et rappelée par Lindelöf (6, p. 103), savoir

$$(s-1)\,\zeta(s) = 4\pi \int_0^\infty (\tfrac{1}{4} + t^2)^{\frac{1}{2}(1-s)} \frac{\cos[(s-1)\operatorname{arctg} 2t]}{(e^{\pi t} + e^{-\pi t})^2}\, dt. \tag{49}$$

Si l'on tient compte, en effet, que

$$b_n = (-1)^{n-1} n \zeta(1-n), \qquad n = 0, 1, 2, \ldots$$

et que

$$(\tfrac{1}{2} + ti)^n + (\tfrac{1}{2} - ti)^n = 2(\tfrac{1}{4} + t^2)^{\frac{1}{2}n} \cos(n \operatorname{arctg} 2t)$$

la formule (49) donne

$$b_n = 2\pi \int_{-\infty}^{+\infty} (-\tfrac{1}{2} + ti)^n \frac{dt}{(e^{\pi t} + e^{-\pi t})^2} \tag{50}$$

qui se ramène immédiatement à (48). Les deux formules (48) et (50) conviennent l'une et l'autre pour exprimer l'orthogonalité (47) des polynômes $Q_n(x)$. En choisissant (48) on aura

$$-\frac{\pi i}{2}\int_{c-i\infty}^{c+i\infty} \frac{Q_m(z)\, Q_n(z)}{\sin^2 \pi z}\, dz = \begin{cases} 0, & m \neq n, \\ 2^n K_n, & m = n, \end{cases}$$

où $-1 < c < 0$ et où K_n a la valeur (46).

Références

1. H. W. Becker and D. H. Browne, *Problem* E461, American Math. Monthly, *48* (1941), 701–703.
2. H. W. Becker and John Riordan, *The arithmetic of Bell and Stirling numbers*, Amer. J. Math. *70* (1948), 385–394.
3. E. T. Bell, *Exponential numbers*, Amer. Math. Monthly, *41* (1934), 411–419.
4. ———, *Exponential polynomials*, Ann. Math. *35* (1934), 258–277.
5. Léo F. Epstein, *A function related to the series for e^{e^x}*, J. Math. Phys. *18* (1939), 153–173.
6. E. Lindelof, *Le Calcul des résidus* (Paris, 1905).

6a. L. Moser and M. Wyman, *An asymptotic formula for the Bell numbers*, Trans. Royal Soc. Can. III, *49* (1955), 49–54.

7. O. Perron, *Die Lehre von den Kettenbrüchen* (Leipzig und Berlin, 1929).
8. John Riordan, *The number of impedances of an n-terminal network*, The Bell System Technical Journal, *18* (1939), 304–314.
9. T. J. Stieltjes, *Oeuvres*, Tome II.
10. G. Szegö, *Orthogonal polynomials* (New York, 1939).
11. J. Touchard, *Sur la fonction Gamma*, Bull. Soc. Math. *41* (1913), 234–242.
12. ———, *Propriétés arithmétiques de certains nombres récurrents*, Ann. Soc. Sci. de Bruzelles, 1933, 21–31.

Lausanne

ON SOME POLYNOMIALS OF TOUCHARD

MAX WYMAN AND LEO MOSER

In the preceding paper Touchard considers a set of polynomials $Q_n(x)$ defined by

$$(1)\quad Q_{n+1}(x) = (2x+1)\,Q_n(x) + \frac{n^4}{4n^2-1}\,Q_{n-1}(x), \quad Q_{-n}(x) = 0, \quad Q_0(x) = 1.$$

Touchard uses (1) to compute $Q_n(x)$ for $0 < n < 9$ and also finds $Q_n(-\frac{1}{2})$. He remarks however "l'expression générale des polynômes $Q_n(x)$ nous echappe." The object of this note is to derive an explicit expression for $Q_n(x)$.

Under the substitution

$$(2)\qquad Q_n = 2^n\binom{2n}{n}^{-1} W_n$$

the conditions (1) become

$$(3)\quad (n+1)W_{n+1} = (2x+1)(2n+1)\,W_n + n^3 W_{n-1}, \quad W_{-n}(x)=0,\ W_0(x)=1.$$

Now define the generating function

$$(4)\qquad W(t) = \sum_{n=0}^{\infty} W_n \frac{t^n}{n!}.$$

The conditions (3) then imply

$$(5)\quad t(t^2-1)\frac{d^2W}{dt^2} + \{3t^2 + 2(2x+1)t - 1\}\frac{dW}{dt} + (t+2x+1)W = 0,$$
$$W(0) = 1.$$

Equation (5) is a special case of Heun's equation. Its solution can be obtained in the following way: Let

$$(6)\qquad W = (1-t)^{-(2x+1)}w, \quad z = t^2.$$

Then (5) becomes

$$(7)\qquad z(z-1)\frac{d^2w}{dz^2} + \{1-(1-2x)z\}\frac{dw}{dz} - x^2w = 0, \qquad w(0) = 1.$$

This is the well-known hypergeometric equation. The only solution regular at $z = 0$ and satisfying the boundary condition is

$$(8)\qquad w = F(-x, -x, 1, +z).$$

Hence from (6) we obtain

$$(9)\qquad W = (1-t)^{-(2x+1)}F(-x, -x, -1, t^2).$$

Received January 11, 1956.

Since

$$W_n = \left.\frac{d^nW}{dt^n}\right|_{t=0}$$

(9) implies

$$(10)\quad W_n = (\Gamma(-(2x+1))\,\Gamma^2(-x))^{-1}\sum_{r=0}^{[\frac{1}{2}n]}\binom{2n}{n}\frac{\Gamma(2x+n-2r+1)\,\Gamma^2(r-x)(2r)!}{(r!)^2}$$
$$= n!\sum_{r=0}^{[\frac{1}{2}n]}\binom{2x+n-2r}{n-2r}\binom{x}{r}^2$$

By (2) and (10) an explicit expression for $Q_n(x)$ is

$$(11)\qquad Q_n(x) = 2^n n!\binom{2n}{n}^{-1}\sum_{r=0}^{[\frac{1}{2}n]}\binom{2x+n-2r}{n-2r}\binom{x}{r}^2.$$

This of course checks with the values of $Q_n(x)$ computed by Touchard for $0 < n < 9$, and also gives his value of $Q_n(-\frac{1}{2})$. Finally, the expression (11) simplifies considerably for x a positive or negative integer and for $2x$ a negative integer. Thus for example

$$(12)\qquad Q_n(0) = 2^n n!\binom{2n}{n}^{-1}$$

and

$$(13)\qquad Q_n(1) = 2^n n!(n^2+n+1)\binom{2n}{n}^{-1}.$$

Equation (13) provides still another simple check on values of $Q_n(x)$ computed from the recurrence formula.

University of Alberta

ON THE THEORY OF RING-LOGICS

ADIL YAQUB

Introduction. Boolean rings $(B, \times, +)$ and Boolean logics (=Boolean algebras) $(B, \cap, ^*)$ are equationally interdefinable in a familiar way **(6)**. Foster's theory of ring-logics **(1; 2; 3)** raises this interdefinability and indeed the entire Boolean theory to a more general level. In this theory a ring (or an algebra) R is studied modulo K, where K is an arbitrary transformation group (or "Coordinate transformations") in R. The Boolean theory results from the special choice, for K, of the "Boolean group," generated by $x^* = 1 - x$ (order 2, $x^{**} = x$). More generally, in a commutative ring $(R, \times, +)$ with identity the *natural group* N, generated by $x^\wedge = 1 + x$ (with $x^\vee = x - 1$ as inverse) was shown to be of particular interest. Thus specialized to N, a commutative ring with identity $(R, \times, +)$ is called a *ring-logic*, mod N, if (1) the $+$ of the ring is equationally definable in terms of its N-logic $(R, \times, ^\wedge, ^\vee)$, and (2) the $+$ of the ring is *fixed* by its N-logic. It was shown **(2)** that each p-ring **(5)** is a ring-logic mod N. It was further shown **(3)** that each p^k-ring **(3; 5)** is a ring-logic mod D, where D is a somewhat more involved group.

All these known examples of ring-logics have zero radical, and the question presents itself: do there exist examples of ring-logics (modulo a suitable group) with non-zero radical? We shall answer this in the affirmative. Indeed, we shall show that the ring of residues mod n (n arbitrary) is a ring-logic modulo the natural group N itself.

1. The ring of residues mod p^k. Let $(R, \times, +)$ be a commutative ring with identity 1. We denote the generator of the natural group N by

$$x^\wedge = 1 + x, \tag{1.1}$$

with inverse

$$x^\vee = x - 1. \tag{1.2}$$

As in **(1)**, we define

$$a \times_\wedge b = (a^\wedge \times b^\wedge)^\vee. \tag{1.3}$$

It is readily verified that

$$a \times_\wedge b = a + b + ab. \tag{1.4}$$

The following notation is used **(2)**:

$$x^{\wedge n} = (\ldots((x^\wedge)^\wedge)\ldots)^\wedge; \quad x^{\vee n} = (\ldots((x^\vee)^\vee)\ldots)^\vee,$$

n iterations. Again

$$x^{\wedge k n} = (x^{\wedge k})^n; \quad x^{\vee k n} = (x^{\vee k})^n.$$

Received September 9, 1955.

We now consider $(R_{p^k}, \times, +)$, the ring of residues mod p^k (p prime) and prove the following

THEOREM 1. $(R_{p^k}, \times, +)$ *is a ring-logic* (mod N). *The ring* $+$ *is given by the following N-logical formula*

$$x + y = \{(x(yx^{p^k-p^{k-1}-1})^{\Delta})x^{p^k-p^{k-1}}\} \times_{\Delta} \{(x^{\Delta}(y(x^{\Delta})^{p^k-p^{k-1}-1})^{\Delta})^{\nabla}(x^{p^k-p^{k-1}})^{\nabla 2}\}. \tag{1.5}$$

Proof. By Euler's generalized form of Fermat's Theorem, we have

$$a^{p^k-p^{k-1}} = 1, \; a \in R_{p^k}, \tag{1.6}$$

a not divisible by p. We now distinguish two cases:

Case 1: Suppose p does not divide x. Then, by (1.6), the right side of (1.5) reduces to

$$\{x(1 + yx^{p^k-p^{k-1}-1})\cdot 1\} \times_{\Delta} 0 = x + yx^{p^k-p^{k-1}} = x + y,$$

since

$$(x^{p^k-p^{k-1}})^{\nabla 2} = 1^{\nabla 2} = 0; \; a \times_{\Delta} 0 = a.$$

This proves (1.5).

Case 2: Suppose p divides x. Then, clearly, p does not divide $x^{\Delta} = 1 + x$. Hence, using Case 1, the right side of (1.5) reduces to

$$\begin{aligned} 0 \times_{\Delta} \{(x^{\Delta}(1 + y(x^{\Delta})^{p^k-p^{k-1}-1}))^{\nabla}\cdot 1\} &= (x^{\Delta} + y(x^{\Delta})^{p^k-p^{k-1}})^{\nabla} \\ &= (x^{\Delta} + y)^{\nabla} = x + y, \end{aligned}$$

since

$$(x^{p^k-p^{k-1}})^{\nabla 2} = 0^{\nabla 2} = 1; \; 0 \times_{\Delta} a = a.$$

Again (1.5) is verified. Hence $(R_{p^k}, \times, +)$ is *equationally* definable in terms of its N-logic. Next, we show that $(R_{p^k}, \times, +)$ is *fixed* by its N-logic.[1] Suppose then that there exists another ring $(R_{p^k}, \times, +')$, with the same class of elements R_{p^k} and the same $\times$ as $(R_{p^k}, \times, +)$ and which has the *same logic* as $(R_{p^k}, \times, +)$. To prove that $+ = +'$. Again we distinguish two cases.

Case 1: p does not divide x. Then

$$x +' y = x(1 +' yx^{p^k-p^{k-1}-1}) = x(yx^{p^k-p^{k-1}-1})^{\Delta} = x(1 + yx^{p^k-p^{k-1}-1}) = x + y,$$

since, by hypothesis, $x^{\Delta} = 1 + x = 1 +' x$.

---

[1]A ring $(R, \times, +)$ is said to be fixed by its N-logic if there exists no other ring $(R, \times, +')$, on the same set R and with the same $\times$ but with $+' \neq +$, which has the same N-logic; i.e.,

$$x^{\Delta} = 1 + x = 1 +' x; \; x^{\nabla} = x - 1 = x -' 1.$$

Case 2: p divides x. Then, clearly, p does not divide $x^{\Delta} = 1 + x$. Hence, by Case 1,

$$x +' y = x^{\Delta} +' y^{\nabla} = x^{\Delta} + y^{\nabla} = x + y.$$

Therefore $+' = +$, and the theorem is proved.

COROLLARY. $(R_p, \times, +) = (F_p, \times, +)$, *the ring (field) of residues* (mod p), *p prime, is a ring-logic* (mod N) *the $+$ being given by setting $k = 1$ in* (1.5), *and making use of $x^p = x$:*

$$(1.7)^2 \qquad x + y = \{(x(x^{p-2}y)^{\Delta})\} \times_{\Delta} \{(x^{\Delta}((x^{\Delta})^{p-2}y)^{\Delta})^{\nabla}(x^{p-1})^{\nabla 2}\}.$$

2. The ring of residues (mod n), n arbitrary. In attempting to generalize Theorem 1 to the residue class ring $(R_n, \times, +)$, where n is *any* positive integer, the following concept of independence, introduced by Foster **(4)**, is needed.

Definition. Let $\mathfrak{A} = \{\mathfrak{A}_1, \mathfrak{A}_2, \ldots, \mathfrak{A}_n\}$ be a finite set of algebras of the same species $\mathfrak{S}$. We say that the algebras $\mathfrak{A}_1, \mathfrak{A}_2, \ldots, \mathfrak{A}_n$ are *independent* if, corresponding to each set $\{\phi_i\}$ of expressions of species $\mathfrak{S}$ $(i = 1, \ldots, n)$, there exists at least one expression X such that

$$\phi_i = X \pmod{\mathfrak{A}_i} \qquad (i = 1, \ldots, n).$$

By an *expression* we mean some composition of one or more indeterminate-symbols $\zeta, \ldots$ in terms of the primitive operations of $\mathfrak{A}_1, \mathfrak{A}_2, \ldots, \mathfrak{A}_n$; $\phi = X$ (mod $\mathfrak{A}$), also written as $\phi = X(\mathfrak{A})$, means that this is an identity of the algebra $\mathfrak{A}$.

We now prove the following

THEOREM 2. *Let $(\mathfrak{A}_1, \times, +), \ldots, (\mathfrak{A}_t, \times, +)$ be a finite set of ring-logics* (mod N), *such that the N-logics $(\mathfrak{A}_1, \times, {}^{\Delta}), \ldots, (\mathfrak{A}_t, \times, {}^{\Delta})$ are independent. Then $\mathfrak{A} = \mathfrak{A}_1 \times \ldots \times \mathfrak{A}_t$ (direct product) is also a ring-logic* (mod N).

Proof. Since $\mathfrak{A}_i$ is a ring-logic (mod N), there exists an N-logical expression ϕ_i such that, for every $x_i, y_i \in \mathfrak{A}_i$ $(i = 1, \ldots, t)$,

$$x_i + y_i = \phi_i = \phi_i(x_i, y_i;\ \times, {}^{\Delta}, {}^{\nabla}) = \phi_i(x_i, y_i;\ \times, {}^{\Delta}).$$

In view of the independence of the logics, there exists an expression X such that

$$X = \begin{cases} \phi_1 \pmod{\mathfrak{A}_1}, \\ \ldots \\ \phi_t \pmod{\mathfrak{A}_t}. \end{cases}$$

Then, for $a = (a_1, a_2, \ldots, a_t) \in \mathfrak{A}$; $b = (b_1, b_2, \ldots, b_t) \in \mathfrak{A}$, we have

[2] This formula is considerably shorter than the formulas for $+$ given in **(2; 3)**.

$$\begin{aligned}X(a, b; \times, {}^{\blacktriangle}) &= X((a_1, a_2, \ldots, a_t), (b_1, b_2, \ldots, b_t); \times, {}^{\blacktriangle})\\ &= (X(a_1, b_1; \times, {}^{\blacktriangle}), X(a_2, b_2; \times, {}^{\blacktriangle}), \ldots, X(a_t, b_t; \times, {}^{\blacktriangle}))\\ &= (a_1 + b_1, a_2 + b_2, \ldots, a_t + b_t)\\ &= (a_1, a_2, \ldots, a_t) + (b_1, b_2, \ldots, b_t)\\ &= a + b;\end{aligned}$$

i.e.,

$$a + b = X(a, b; \times, {}^{\blacktriangle}); a, b \in \mathfrak{A}.$$

Hence, $\mathfrak{A} = \mathfrak{A}_1 \times \ldots \times \mathfrak{A}_t$ is *equationally* definable in terms of its N-logic. Next, we show that $\mathfrak{A}$ is fixed by its N-logic. Suppose there exists a $+'$ such that $(\mathfrak{A}, \times, +')$ is a ring, with the same class of elements $\mathfrak{A}$ and the same $\times$ as the ring $(\mathfrak{A}, \times, +)$, and which has the *same logic* $(\mathfrak{A}, \times, {}^{\blacktriangle})$ as the ring $(\mathfrak{A}, \times, +)$. To prove that $+ = +'$.

Now, let $a = (a_1, a_2, \ldots, a_t) \in \mathfrak{A}$; $b = (b_1, b_2, \ldots, b_t) \in \mathfrak{A}$. A new $+'$ in $\mathfrak{A}$ defines and is defined by new $+'_1$ in $\mathfrak{A}_1$, $+'_2$ in $\mathfrak{A}_2, \ldots, +'_t$ in $\mathfrak{A}_t$, such that $(\mathfrak{A}_1, \times, +'_1)$ is a ring, and similarly for $(\mathfrak{A}_2, \times, +'_2), \ldots, (\mathfrak{A}_t, \times, +'_t)$; i.e.,

$$\begin{aligned}a +' b &= (a_1, a_2, \ldots, a_t) +' (b_1, b_2, \ldots, b_t)\\ &= (a_1 +'_1 b_1, a_2 +'_2 b_2, \ldots, a_t +'_t b_t).\end{aligned} \tag{2.1}$$

Furthermore, the assumption that $(\mathfrak{A}, \times, +')$ has the same logic as $(\mathfrak{A}, \times, +)$ is equivalent to the assumption that $(\mathfrak{A}_1, \times, +'_1)$ has the same logic as $(\mathfrak{A}_1, \times, +)$, and similarly for $(\mathfrak{A}_i, \times, +'_i)$ and $(\mathfrak{A}_i, \times, +) (i = 2, \ldots, t)$. Since $(\mathfrak{A}_1, \times, +)$ is a ring-logic, and hence with its $+$ fixed, it follows that $+'_1 = +$; similarly $+'_2 = +, \ldots, +'_t = +$. Hence, using (2.1), $+' = +$, and the proof is complete.

We shall now prove the following

LEMMA 3. *Let $p_1, \ldots, p_t$ be distinct primes, and let*

$$(R_{n_i}, \times, +),\ n_i = p_i^{k_i} = p_i m_i;\ i = 1, \ldots, t,$$

be a set of residue class rings (mod n_i). *Then the logics $(R_{n_i}, \times, {}^{\blacktriangle}) (i = 1, \ldots, t)$ are independent.*

Proof. Let

$$P(i) = \prod_{j=1}^{t} n_j, \qquad j \neq i,$$

Then, clearly

$$(P(i), n_i) = 1.$$

Hence, there exist integers $r_i > 0$, $s_i > 0$ such that

$$r_i P(i) - s_i n_i = 1.$$

Now, define

$$\epsilon(x) = x^{(n_1 - m_1)(n_2 - m_2) \ldots (n_t - m_t)}.$$

Then one easily verifies that, for $i \neq j$,

$$\omega_i = \omega_i(x) = \{\epsilon(x) \times_{\Delta}((\epsilon(x))^{\nabla})^{(n_1-m_1)\dots(n_t-m_t)}\}^{\Delta r_i P(t)-1} = \begin{cases} 1(R_{n_i}) \\ 0(R_{n_j}) \end{cases}.$$

Now, to prove the independence of the logics $(R_{n_i}, \times, ^{\Delta})$, let $\{\phi_i\}$ be a set of t expressions of species $\times, ^{\Delta}$; i.e., a primitive composition of indeterminate-symbols in terms of the operations $\times, ^{\Delta}$; then, if we define (cf. **4**)

$$X = \phi_1\omega_1 \times_{\Delta}\phi_2\omega_2 \times_{\Delta} \dots \times_{\Delta}\phi_t\omega_t,$$

we immediately obtain

$$\phi_i = X(\text{mod } R_{n_i}),$$

since $a \times_{\Delta} 0 = a = 0 \times_{\Delta} a$. This proves the theorem.

Recalling the well-known fact that

$$(2.2) \qquad (R_n, \times, +) \cong R_{n_1} \times \dots \times R_{n_t} \text{ (direct product)},$$

n arbitrary, $n = n_1 \dots n_t$, a combination of Theorems 1, 2, Lemma 3 and (2.2) readily yields

THEOREM 4 (Fundamental Theorem on R_n as ring-logics). $(R_n, \times, +)$, *the residue class ring* (mod n), *n arbitrary, is a ring-logic* (mod N).

We conclude with several illustrative examples.

Example 1. $R_{p^k} = R_2 = F_2 = \{0, 1\}$.
It is readily verified that each of (1.5) and (1.7) reduces to the familiar Boolean formula

$$(2.3) \qquad x + y = xy^{\Delta} \times_{\Delta} x^{\Delta}y.$$

Example 2. $R_{p^k} = R_3 = F_3 = \{0, 1, 2\}$.
Formula (1.7) yields

$$(2.4) \qquad x + y = \{x(xy)^{\Delta}\} \times_{\Delta}\{[(x^{\Delta}(x^{\Delta}y)^{\Delta})]^{\nabla}(x^2)^{\nabla 2}\}.$$

Compare with **(1)** in which the following formula was obtained:

$$(2.5) \qquad x + y = xy^{\Delta} \times_{\Delta} x^{\Delta}y \times_{\Delta} x^2y^2.$$

It is noteworthy to observe that the $+$ of $(F_p, \times, +)$, the field of residues (mod p), p prime, may also be expressed in the following form:

$$(2.6) \qquad x + y = \{x(yx^{p-2})^{\Delta}\} \times_{\Delta}\{y(x^{\Delta}x^{\Delta^2} \dots x^{\Delta^{p-1}})^2\}.$$

or by

$$(2.7) \qquad x + y = \{x(yx^{p-2})^{\Delta}\} \times_{\Delta}\{y(x^{p-1})^{\nabla 2}\}.$$

The last formula, when specialized to F_3, gives a simpler expression for $+$ than (2.4).

Example 3. $R_{p^k} = R_{2^2} = \{0, 1, 2, 3\}$.
Formula (1.5) reduces to

$$x + y = \{(x(xy)^{\Delta}x^2)\} \times_{\Delta} \{[(x^{\Delta}(x^{\Delta}y)^{\Delta})]^{\nabla}(x^2)^{\nabla 2}\}. \tag{2.8}$$

It may be verified that the $+$ in $(R_4, \times, +)$ is also given by

$$x + y = \{(xy)^{\Delta}(xy)^2 \times_{\Delta} (x \times_{\Delta} y)(xy)^{\Delta 2}\}\{(xy)(xy)^{2\nabla}\}^{\Delta}. \tag{2.9}$$

This last formula excels most of the others in obviously displaying the symmetry of $+$.

Example 4. $R_n = R_6 = \{0, 1, 2, 3, 4, 5\}$.
The correspondence

$$\begin{array}{ll} 0 \leftrightarrow (0_2, 0_3), & 3 \leftrightarrow (1_2, 0_3), \\ 1 \leftrightarrow (1_2, 1_3), & 4 \leftrightarrow (0_2, 1_3), \\ 2 \leftrightarrow (0_2, 2_3), & 5 \leftrightarrow (1_2, 2_3), \end{array}$$

determines an isomorphism of R_6 and $R_2 \times R_3$ (direct product), where $R_2 = \{0_2, 1_2\}$ and $R_3 = \{0_3, 1_3, 2_3\}$.

It is readily verified (compare with the proof of Lemma 3 and (2.3), (2.5) above) that

$$\begin{aligned} x + y = \{&(xy^{\Delta} \times_{\Delta} x^{\Delta}y)(x^2 \times_{\Delta} (x^2)^{\nabla 2})^{\Delta 2}\} \\ &\times_{\Delta} \{(xy^{\Delta} \times_{\Delta} x^{\Delta}y \times_{\Delta} x^2y^2)(x^2 \times_{\Delta} (x^2)^{\nabla 2})^{\Delta 2}\}. \end{aligned} \tag{2.10}$$

Formula (2.10) may be verified either by direct substitution from R_6, or via the $R_2 \times R_3$ representation above.

References

1. A. L. Foster, *On n-ality theories in rings and their logical algebras, including tri-ality principle in three-valued logics*, Amer. J. Math., *72* (1950), 101–123.
2. ———, *p-rings and ring-logics*, Univ. Calif. Publ., *1* (1951), 385–396.
3. ———, *p^k-rings and ring-logics*, Ann. Scu. Norm. Pisa, *5* (1951), 279–300.
4. ———, *Unique subdirect factorization within certain classes of universal algebras*, Math. Z., *62* (1955), 171–188.
5. N. H. McCoy and D. Montgomery, *A representation of generalized Boolean rings*, Duke Math. J., *3* (1937), 455–459.
6. M. H. Stone, *The theory of representations of Boolean algebras*, Trans. Amer. Math. Soc., *40* (1936), 37–111.

University of California, Berkeley
and
Purdue University

MASCHKE MODULES OVER DEDEKIND RINGS

IRVING REINER

1. Introduction. We use the following notation throughout:

$\mathfrak{o}$ = Dedekind ring (**8**; **12**, p. 83).
K = quotient field of $\mathfrak{o}$.
A = finite-dimensional separable algebra over K, with identity element e (**6**, p. 115).
G = $\mathfrak{o}$-order in A (**2**, p. 69).
$\mathfrak{p}$ = prime ideal in $\mathfrak{o}$.
$K_{\mathfrak{p}}$ = $\mathfrak{p}$-adic completion of K.
$\mathfrak{o}_{\mathfrak{p}}$ = $\mathfrak{p}$-adic integers in $K_{\mathfrak{p}}$.
$\mathfrak{p}^* = \pi\mathfrak{o}_{\mathfrak{p}}$ = unique prime ideal in $\mathfrak{o}_{\mathfrak{p}}$.
$\bar{K} = \mathfrak{o}/\mathfrak{p} = \mathfrak{o}_{\mathfrak{p}}/\mathfrak{p}^*$ = residue class field.

By a *G-module* we shall mean a left G-module R satisfying

1. R is a finitely generated torsion-free left $\mathfrak{o}$-module.
2. For $x, y \in G, r, s \in R$:

$$(xy)r = x(yr), \quad (x+y)r = xr + yr, \quad x(r+s) = xr + xs, \quad er = r.$$

Following Gaschütz and Ikeda (**3**; **5**; see also **7**; **10**) we call a G-module R an *M_u-G-module* (unterer Maschke Modul) if, whenever R is an $\mathfrak{o}$-direct summand of a G-module S, R is a G-direct summand of S. Likewise, R is an *M_0-G-module* (oberer Maschke Modul) if, whenever S/R_1 is G-isomorphic to R where the G-module S contains the G-module R_1 as $\mathfrak{o}$-direct summand, R_1 is a G-direct summand of S.

If all modules considered happen to have $\mathfrak{o}$-bases (for example, when $\mathfrak{o}$ is a principal ideal ring), then we may interpret these concepts in terms of matrix representations over $\mathfrak{o}$. Thus, a representation Γ of G in $\mathfrak{o}$ is an *M_0-representation* if for every reduced representation

$$\begin{pmatrix} \Gamma & \Lambda \\ 0 & \Delta \end{pmatrix}$$

of G in $\mathfrak{o}$, the binding system Λ is strongly-equivalent (**13**) to zero, that is, there exists a matrix T (over $\mathfrak{o}$) such that

$$\Lambda(x) = \Gamma(x)T - T\Delta(x) \qquad \text{for all } x \in G.$$

(Likewise we may define an M_u-representation of G in $\mathfrak{o}$.)

Received September 19, 1955; in revised form December 14, 1955. This work was supported in part by a contract with the National Science Foundation. The author wishes to thank Dr. P. Roquette for some helpful conversations during the preparation of this paper.

Starting with a prime ideal $\mathfrak{p}$ of $\mathfrak{o}$, we may form $\bar{G} = G/\mathfrak{p}G$, an algebra over $\bar{K}$. If R is a G-module, then $\bar{R} = R/\mathfrak{p}R$ can be made into a $\bar{G}$-module in obvious fashion, and $\bar{R}$ is then a vector space over $\bar{K}$. The main results of this note are as follows:

THEOREM 1. *If for each $\mathfrak{p}$, $\bar{R}$ is an M_u-$\bar{G}$-module (or M_0-$\bar{G}$-module), then R is an M_u-G-module (or M_0-G-module).*

THEOREM 2. *If G is a Frobenius algebra over $\mathfrak{o}$, and R is an M_u-G-module (or M_0-G-module), then for each $\mathfrak{p}$, $\bar{R}$ is an M_u-$\bar{G}$-module (or M_0-$\bar{G}$-module).*

The significance of Theorem 1 is that it reduces the problem of deciding whether an $\mathfrak{o}$-module R is an M_u-G-module to that of determining for each $\mathfrak{p}$ whether the vector space $\bar{R}$ over $\bar{K}$ is an M_u-$\bar{G}$-module. Thus, we pass from a *ring* problem to a *field* problem, which is in general much simpler.

In the important case where $G = \mathfrak{o}(H)$ is the group ring of a finite group H, then $\bar{G}$ is semi-simple whenever $\mathfrak{p}$ does not divide the order of H, and for such $\mathfrak{p}$ the module $\bar{R}$ is automatically an M-$\bar{G}$-module. More generally, we may form the ideal $I(G)$ of G defined by Higman **(4)**; his results show that $I(G) \neq 0$ in this case. From **(9)** we deduce at once that $\bar{G}$ is semi-simple whenever $\mathfrak{p}$ does not divide $I(G)$. Therefore:

COROLLARY 1. *R is an M_u-G-module (or M_0-G-module) if for each $\mathfrak{p}$ dividing $I(G)$, $\bar{R}$ is an M_u-$\bar{G}$-module (or M_0-$\bar{G}$-module).* (Note that only finitely many $\mathfrak{p}$'s are involved.)

Now let G be a Frobenius algebra over $\mathfrak{o}$, for example, $G = \mathfrak{o}(H)$. Then by **(5)** there is no distinction between M_0- and M_u-modules, and Theorems 1 and 2 tell us that R is an M-G-module if and only if for each $\mathfrak{p}$, $\bar{R}$ is an M-$\bar{G}$-module. Using the concept of *genus* introduced by Maranda in **(9)**, we have:

COROLLARY 2. *Let G be a Frobenius algebra over $\mathfrak{o}$, and let R, S be G-modules in the same genus. Then R is an M-G-module if and only if S is an M-G-module.*

2. $\mathfrak{p}$-adic completion. Theorem 1 will follow at once from two lemmas, of which we prove the more difficult first. Let R be a G-module, and define

$$G_{\mathfrak{p}} = G \otimes \mathfrak{o}_{\mathfrak{p}}, \quad R_{\mathfrak{p}} = \mathfrak{o}_{\mathfrak{p}} \otimes R,$$

both products being taken over $\mathfrak{o}$.

LEMMA 1. *If for each $\mathfrak{p}$, $R_{\mathfrak{p}}$ is an M_u-$G_{\mathfrak{p}}$-module (or M_0-$G_{\mathfrak{p}}$-module), then R is an M_u-G-module (or M_0-G-module).*

Proof. (We give the proof only for M_u-modules.) Let R be an $\mathfrak{o}$-direct summand of a G-module S. We wish to show that R is a G-direct summand of S, that is, that there exists $f \in \mathrm{Hom}_G(S, R)$ such that $f|R =$ identity. Using

the Steinitz-Chevalley theory **(1; 11)** of the structure of finitely generated torsion-free modules over Dedekind rings, and taking into account the hypothesis that R is an $\mathfrak{o}$-direct summand of S, we may write

$$S = \mathfrak{A}_1 s_1 \oplus \ldots \oplus \mathfrak{A}_n s_n, \quad R = \mathfrak{A}_1 s_1 \oplus \ldots \oplus \mathfrak{A}_m s_m,$$

with $m < n$, where each $\mathfrak{A}_i$ is an $\mathfrak{o}$-ideal in K, and where $s_1, \ldots, s_n$ are linearly independent over K. For the remainder of this proof, let the index i range from 1 to n, and j from 1 to m.

To prove the lemma, it suffices to exhibit $f \in \mathrm{Hom}_A(KS, KR)$ such that $f|KR$ = identity, and f maps S into R. (We use KS to denote the K-module generated by S.) Let us set

$$f(s_i) = \sum a_{ij} s_j, \quad a_{ij} \in K, \tag{1}$$

thereby defining $f \in \mathrm{Hom}_K(KS, KR)$. Then f maps S into R if and only if for each $\alpha \in \mathfrak{A}_i$ we have $\alpha a_{ij} \in \mathfrak{A}_j$, that is, if and only if

$$a_{ij} \in (\mathfrak{A}_j : \mathfrak{A}_i) \qquad \text{for all } i, j. \tag{2}$$

On the other hand, the map f defined by (1) will be an A-homomorphism with $f|KR$ = identity, if and only if for all $x \in G$, $s \in S$, $r \in R$:

$$f(xs) = xf(s), \quad f(r) = r.$$

Let us set

$$G = \mathfrak{o}x_1 + \ldots + \mathfrak{o}x_t.$$

This is possible since **(2**, p. 70**)** G is a finitely generated $\mathfrak{o}$-module. Then f is an A-homomorphism with $f|KR$ = identity, if and only if

$$f(x_k s_i) = x_k f(s_i), \quad f(s_j) = s_j \qquad \text{for all } i, j, k, \tag{3}$$

where the index k ranges from 1 to t. Equations (3) are a set of linear equations with coefficients in K, to be solved for unknowns $\{a_{ij}\}$ satisfying (2).

From the hypotheses of the lemma we deduce that for each $\mathfrak{p}$, (3) has a solution $\{a_{ij}\}$ satisfying $a_{ij} \in (\mathfrak{A}_j : \mathfrak{A}_i)\mathfrak{o}_\mathfrak{p}$ for all i, j. Thus (3) is solvable over the extension field $K_\mathfrak{p}$ of K, and hence is also solvable over K. The general solution of (3) over K is given by

$$a_{ij} = e_{ij}/d_{ij}, \; e_{ij} = e_{ij}(\mathfrak{t}) = b_{ij} + \sum_{\nu=1}^{N} c_{ij}^{(\nu)} t_\nu, \tag{4}$$

where the $b_{ij}, c_{ij}^{(\nu)}, d_{ij}$ are fixed elements of $\mathfrak{o}$, $d_{ij} \neq 0$, and where $\mathfrak{t}$ ranges over all N-tuples in K^N. The general solution of (3) over $K_\mathfrak{p}$ is also given by (4) by letting $\mathfrak{t}$ range over $K_\mathfrak{p}^N$. Then for each $\mathfrak{p}$, we can find $\mathfrak{t}(\mathfrak{p})$ for which

$$e_{ij}(\mathfrak{t}(\mathfrak{p})) \in \mathfrak{B}_{ij}\mathfrak{o}_\mathfrak{p} \qquad \text{for all } i, j, \tag{5}$$

where $\mathfrak{B}_{ij} = (\mathfrak{A}_j : \mathfrak{A}_i) d_{ij}$.

For each $\mathfrak{p}$, let $b(\mathfrak{p})$ be the maximal exponent to which $\mathfrak{p}$ occurs in the prime ideal factorizations of the ideals $\mathfrak{B}_{ij}$. Then $b(\mathfrak{p}) = 0$ except for a finite set of primes. Set $P = \{\mathfrak{p}: b(\mathfrak{p}) > 0\}$, and choose an N-tuple $\mathfrak{t}$ with components in $\mathfrak{o}$ such that (componentwise)

$$\mathfrak{t} \equiv \mathfrak{t}(\mathfrak{p}) \pmod{\mathfrak{p}^{b(\mathfrak{p})}} \qquad \text{for each } \mathfrak{p} \in P.$$

In that case, $e_{ij}(\mathfrak{t}) \equiv e_{ij}(\mathfrak{t}(\mathfrak{p})) \pmod{\mathfrak{p}^{b(\mathfrak{p})}}$ for each $\mathfrak{p} \in P$, and all i, j, whence by (5) we have

$$\text{(6)} \qquad \operatorname{ord}_{\mathfrak{p}} e_{ij}(\mathfrak{t}) \geqslant \operatorname{ord}_{\mathfrak{p}} \mathfrak{B}_{ij} \qquad \text{for all } i, j,$$

for all $\mathfrak{p} \in P$. But for $\mathfrak{p} \notin P$, equation (6) is certainly valid because $e_{ij}(\mathfrak{t}) \in \mathfrak{o}$, and $\operatorname{ord}_{\mathfrak{p}} \mathfrak{B}_{ij} \leqslant 0$. Hence we deduce that $e_{ij}(\mathfrak{t}) \in \mathfrak{B}_{ij} = (\mathfrak{A}_j : \mathfrak{A}_i) d_{ij}$ for all i, j, whence (4) gives a solution of (3) for which (2) holds.

We may remark that this lemma is almost trivial when $\mathfrak{o}$ is a principal ideal ring.

3. Modular representations. Now let $R_{\mathfrak{p}}$ be a $G_{\mathfrak{p}}$-module, and define $\bar{R}_{\mathfrak{p}} = R_{\mathfrak{p}}/\pi R_{\mathfrak{p}}$, $\bar{G}_{\mathfrak{p}} = G_{\mathfrak{p}}/\pi G_{\mathfrak{p}}$. To complete the proof of Theorem 1, we need only show:

LEMMA 2. *If $\bar{R}_{\mathfrak{p}}$ is an M_u-$\bar{G}_{\mathfrak{p}}$-module (or M_0-$\bar{G}_{\mathfrak{p}}$-module), then $R_{\mathfrak{p}}$ is an M_u-$G_{\mathfrak{p}}$-module (or M_0-$G_{\mathfrak{p}}$-module).*

Proof. Since $\mathfrak{o}_{\mathfrak{p}}$ is a principal ideal ring, we may express the proof (given here only for M_0-modules) in terms of matrix representations. We must show that if Γ is a representation of $G_{\mathfrak{p}}$ in $\mathfrak{o}_{\mathfrak{p}}$ for which $\bar{\Gamma}$ (the induced modular representation of $\bar{G}_{\mathfrak{p}}$ in $\bar{K}$) is an M_0-representation, then in any reduced representation

$$\text{(7)} \qquad \begin{pmatrix} \Gamma & \Lambda \\ 0 & \Delta \end{pmatrix}$$

of $G_{\mathfrak{p}}$ in $\mathfrak{o}_{\mathfrak{p}}$, the binding system Λ is strongly-equivalent to zero.

We may write $G_{\mathfrak{p}} = \mathfrak{o}_{\mathfrak{p}} y_1 \oplus \ldots \oplus \mathfrak{o}_{\mathfrak{p}} y_n$, $\bar{G}_{\mathfrak{p}} = \bar{K} y_1 \oplus \ldots \oplus \bar{K} y_n$. We shall show the existence of a matrix T over $\mathfrak{o}_{\mathfrak{p}}$ such that

$$\text{(8)} \qquad \Lambda(y_i) = \Gamma(y_i) T - T \Delta(y_i) \qquad \text{for each } i,$$

where in this proof the index i ranges from 1 to n. By taking residue classes mod $\mathfrak{p}^*$, the representation (7) gives a representation

$$\begin{pmatrix} \bar{\Gamma} & \bar{\Lambda} \\ 0 & \bar{\Delta} \end{pmatrix}$$

of $\bar{G}_{\mathfrak{p}}$ in $\bar{K}$. Since $\bar{\Gamma}$ is by hypothesis an M_0-representation, the binding system $\bar{\Lambda}$ is strongly-equivalent to zero over $\bar{K}$. Therefore there exists V_1 over $\mathfrak{o}_{\mathfrak{p}}$ such that

$$\text{(9)} \qquad \Lambda(y_i) = \Gamma(y_i) V_1 - V_1 \Delta(y_i) + \pi \Lambda^{(1)}(y_i) \qquad \text{for each } i,$$

where $\Lambda^{(1)}$ is also over $\mathfrak{o}_\mathfrak{p}$. But then (7) with Λ replaced by $\Lambda^{(1)}$ gives another $\mathfrak{o}_\mathfrak{p}$-representation of $G_\mathfrak{p}$, whence the same argument shows

$$\Lambda^{(1)}(y_i) = \Gamma(y_i)\, V_2 - V_2\Delta(y_i) + \pi\,\Lambda^{(2)}(y_i) \qquad \text{for all } i,$$

where V_2 and $\Lambda^{(2)}$ are over $\mathfrak{o}_\mathfrak{p}$. Continuing in this way, we obtain a solution of (8) given by $T = V_1 + \pi V_2 + \pi^2 V_3 + \ldots$.

This proof could also have been stated in terms of cohomology groups.

4. Frobenius algebra. Suppose in this section that G is a Frobenius algebra over $\mathfrak{o}$, that is, there exist $\mathfrak{o}$-bases $\{u_i\}$, $\{v_i\}$ of G (called *dual* bases) such that the right regular representation of G with respect to $\{v_i\}$ coincides with the left regular representation with respect to $\{u_i\}$. Assume that G has an $\mathfrak{o}$-basis containing e. Ikeda showed **(5)** that M_0- and M_u-modules were the same, and that a G-module R is an M-G-module if and only if there exists an $\mathfrak{o}$-endomorphism ϕ of R such that

$$(10) \qquad \sum u_i \phi v_i = \text{identity endomorphism of } R.$$

Gaschütz **(3)** had shown this for the case where $G = \mathfrak{o}(H)$, H = finite group, with (10) replaced by:

$$(11) \qquad \sum_{h \in H} h\, \phi\, h^{-1} = \text{identity endomorphism of } R.$$

We may use Ikeda's result to obtain an immediate proof of Theorem 2. By hypothesis, R is an M-G-module, whence (10) holds for some $\mathfrak{o}$-endomorphism ϕ. But then clearly ϕ induces a $\bar{K}$-endomorphism $\bar{\phi}$ of $\bar{R}$, and $\sum u_i \phi v_i$ = identity endomorphism of $\bar{R}$, so that $\bar{R}$ is an M-$\bar{G}$-module.

References

1. C. Chevalley, *L'arithmétique dans les algèbres de matrices*, Act. Sci. et Ind. *229* (1935).
2. M. Deuring, *Algebren* (Berlin, 1949).
3. W. Gaschütz, *Ueber den Fundamentalsatz von Maschke zur Darstellungstheorie der endlichen Gruppen*, Math. Z., *56* (1952), 376–387.
4. D. G. Higman, *On orders in separable algebras*, Can. J. Math., *7* (1955), 509–515.
5. M. Ikeda, *On a theorem of Gaschütz*, Osaka Math. J., *5* (1953), 53–58.
6. N. Jacobson, *The Theory of Rings* (New York, 1943).
7. F. Kasch, *Grundlagen einer Theorie der Frobeniuserweiterungen*, Math. Ann., *127* (1954), 453–474.
8. I. Kaplansky, *Modules over Dedekind rings and valuation rings*, Trans. Amer. Math. Soc., *72* (1952), 327–340.
9. J.-M. Maranda, *On the equivalence of representations of finite groups by groups of automorphisms of modules over Dedekind rings*, Can. J. Math., *7* (1955), 516–526.
10. H. Nagao and T. Nakayama, *On the structure of* (M_0)- *and* (M_u)-*modules*, Math. Z., *59* (1953), 164–170.

11. E. Steinitz, *Rechteckige Systeme und Moduln in algebraischen Zahlkörpern*, Math. Ann. I, *71* (1911), 328–354; II, *72* (1912), 297–345.
12. B. L. van der Waerden, *Modern Algebra*, II (New York, 1950).
13. H. Zassenhaus, *Neuer Beweis der Endlichkeit der Klassenzahl bei unimodularer Äquivalenz endlicher ganzzahliger Substitutionsgruppen*, Hamb. Abh., *12* (1938), 276–288.

Institute for Advanced Study and
University of Illinois

STANDARD AND ACCESSIBLE RINGS

ERWIN KLEINFELD

1. Introduction. A ring is defined to be standard **(1)** in case the following two identities hold:

$$(wx, y, z) + (xz, y, w) + (wz, y, x) = 0, \tag{1}$$

$$(x, y, z) + (z, x, y) - (x, z, y) = 0, \tag{2}$$

where the associator (x, y, z) is defined by $(x, y, z) = (xy)z - x(yz)$. Albert has determined the structure of finite-dimensional, standard algebras **(1)**. The simple ones turn out to be either Jordan algebras or associative ones.

We focus attention here on a more general class of rings, which we shall call accessible. By permuting w and x in (1) and subtracting from (1) we obtain the identity

$$((w, x), y, z) = 0, \tag{3}$$

where the commutator (w, x) is defined by $(w, x) = wx - xw$. A ring is called accessible in case identities (2) and (3) hold. Thus a standard ring is automatically accessible. On the other hand, while (2) and (3) hold in any commutative ring, (1) need not.

The structure of accessible rings, without finiteness assumptions, can readily be determined. An accessible ring is defined to be simple in case it has no proper two-sided ideals. Simple, accessible rings are either associative or commutative. From this result it follows trivially that simple, standard rings of characteristic different from 3 are either Jordan or associative rings. A structure for semi-simple, accessible rings is given, utilizing the Jacobson-Brown radical and the fact that primitive, accessible rings are either associative or commutative.

The following result may also be of interest. If an accessible ring has no nilpotent ideals other than zero, then it is isomorphic to a subdirect sum of an associative and a commutative ring. Hence all identities common to the class of rings, which consist of all associative rings and all commutative rings, must hold in such a ring.

The methods of proof are quite elementary. Identities are obtained which enable the construction of certain significant ideals.

2. Preliminaries. Substituting $z = y$ in (2) one obtains the flexible law, $(y, x, y) = 0$. A linearization of this identity yields $(y, x, z) = -(z, x, y)$.

Received September 26, 1955. This research was supported in part by a grant from the National Science Foundation to the University of Connecticut.

As Albert observed, it can now be seen that (2) is equivalent to the flexible law and the identity

$$(x, y, z) + (y, z, x) + (z, x, y) = 0.$$

We note that similarly (1) is equivalent to (3) and the identity

$$(wx, y, z) + (xz, y, w) + (zw, y, x) = 0.$$

In an arbitrary ring the identity

$$(xy, z) = x(y, z) + (x, z)y + (x, y, z) + (z, x, y) - (x, z, y)$$

holds. Thus (2) is a consequence of the commutative law, as well as of the associative law. Moreover the identity

$$(xy, z) = x(y, z) + (x, z)y, \tag{4}$$

holds in every accessible ring.

Another identity which holds in an arbitrary ring is

$$(wx, y, z) - (w, xy, z) + (w, x, yz) = w(x, y, z) + (w, x, y)z. \tag{5}$$

The nucleus of an accessible ring is defined as the set of all elements n in R with the property $(n, R, R) = 0$. If n is an element of the nucleus N of R, then because of the flexible law $(R, R, n) = 0$. Finally, because of (2), it follows that also $(R, n, R) = 0$. If n is substituted for w in (5), it becomes obvious that

$$(nx, y, z) = n(x, y, z), \qquad n \in N. \tag{6}$$

The center C of R is defined as the set of all elements c in N which have the additional property that $(c, R) = 0$.

We now proceed to develop further identities that hold in arbitrary accessible rings. The elements u, v, w, x, y, z will denote arbitrary elements of such rings.

Through repeated use of (4) one may break up $((w, x, y), z)$ as

$$\begin{aligned}((w, x, y), z) &= (wx\cdot y - w\cdot xy, z) = wx\cdot(y, z) + w(x, z)\cdot y \\ &\quad + (w, z)x\cdot y - (w, z)\cdot xy - w\cdot x(y, z) - w\cdot(x, z)y \\ &= (w, x, (y, z)) + (w, (x, z), y) + ((w, z), x, y).\end{aligned}$$

Since (3) implies that every commutator is in the nucleus, we obtain

$$((w, x, y), z) = 0. \tag{7}$$

Because of (6) and the fact that every commutator is in the nucleus we get $(v, x)(x, y, z) = ((v, x)x, y, z)$. It follows from (4) that $(v, x)x = (vx, x)$. Consequently

$$((v, x)x, y, z) = ((vx, x), y, z) = 0.$$

Therefore $(v, x)(x, y, z) = 0$.[1] A linearization of this last identity becomes

(8) $$(v, w)(x, y, z) = -(v, x)(w, y, z).$$

One can now prove that a product of a commutator with an associator always lies in the center. First one notes that

$$((v, w)(x, y, z), u) = ((v, w), u)(x, y, z),$$

because of (4) and (7). From the definition of the commutator it follows that

$$((v, w), u)(x, y, z) = -(u, (v, w))(x, y, z).$$

It is this last form to which we apply (8) to obtain

$$-(u, (v, w))(x, y, z) = (u, x)((v, w), y, z).$$

Finally (3) tells us that $((v, w), y, z) = 0$, so that $(u, x)((v, w), y, z) = 0$. Consequently $((v, w)(x, y, z), u) = 0$. It remains only to prove that $(v, w)(x, y, z)$ lies in the nucleus. It is easily seen that

$$((v, w)(x, y, z), t, u) = (v, w)((x, y, z), t, u),$$

using (6) and (3). At this point (8) is employed to yield

$$(v, w)((x, y, z), t, u) = -(v,(x, y, z))(w, t, u).$$

But $(v,(x, y, z))=0$ was proven with (7). Consequently $((v, w)(x, y, z), t, u)=0$. We have established that

(9) $$(v, w)(x, y, z) \in C.$$

Now let us consider the element $[(v, w)(x, y, z)]^2$. Clearly

$$[(v, w)(x, y, z)]^2 = (v, w)(x, y, z)(v, w)(x, y, z) = -(v, x)(w, y, z)(v, w)(x, y, z),$$

using (3) and (8). On the other hand $(w, y, z)(v, w) = (v, w)(w, y, z)$, because of (7). But we have already noted that $(v, w)(w, y, z) = 0$. Thus we have proved that

(10) $$[(v, w)(x, y, z)]^2 = 0.$$

3. Structure theory. Decently behaved rings have no nilpotent elements in their center. For let R be any ring with nilpotent elements in its center. Then there must be an element $c \neq 0$ and in the center of R such that $c^2 = 0$. Consider the ideal D generated by c. It consists of all elements of the form $ic + cx$, where i is any integer and x an arbitrary element of R. It is now easy to verify that $D^2 = 0$, and so R has a non-zero, nilpotent ideal.

Henceforth we shall be considering accessible rings R without nilpotent

---

[1]Independently R. L. San Soucie has announced in Abstract 672, Bull. A. M. S. *61* (1955) that rings satisfying (3), which have no divisors of zero, are either associative or commutative.

elements in their centers, unless otherwise noted. An immediate consequence of this assumption, taking into account (9) and (10), is that

$$(v, w)(x, y, z) = 0. \tag{11}$$

But then one can also obtain from (4) and (11) that

$$(v(x, y, z), w) = (v, w)(x, y, z) = 0.$$

Also

$$((v, w)x, y, z) = (v, w)(x, y, z) = 0,$$

because of (6) and (11). This last information allows us to construct ideals A and B in R, which have rather interesting properties. Let A consist of all finite sums of elements of the form (x, y, z) or of the form $w(x, y, z)$. This set A, as may be readily verified, is an ideal even in an arbitrary ring. It is the smallest ideal modulo which the ring is associative. With the present assumptions, namely that R is accessible and has no nilpotent elements in its center, we can assert that for any element a in A we have $(a, R) = 0$.

Let B consist of all finite sums of elements of the form (x, y) or of the form $(x, y)z$. In an arbitrary ring this set need not be an ideal, but by virtue of (3) and (4) it can be shown to be one. In addition it is also true that B is contained in the nucleus N. B is also the smallest ideal modulo which R is commutative.

From previous remarks, in conjunction with (7) and (11), it becomes clear that for any element a in A and any element b in B we must have $ab = 0$. Therefore $AB = 0$. Suppose that x is an element of $A \cap B$. Then since $AB = 0$, $x^2 = 0$. But x lies in the center of R because of the previously mentioned properties of A and B. Hence $x = 0$.

At this point several theorems may be established.

THEOREM 1. *A simple, accessible ring R is either associative or commutative.*

Proof. If R has nilpotent elements in its center then the ideal D described previously is different from zero, so that $D = R$. Since $D^2 = 0$, R must be a trivial ring, which is both associative and commutative. The only remaining case is the one in which R has no nilpotent elements in its center. Then the ideal B constructed above is either zero or the whole ring. If $B = 0$ then R is commutative, while if $B = R$ then R is associative, since B is contained in the nucleus. This completes the proof.

By substituting $w = x$ and $z = x$ in (1), one obtains $3(x^2, y, x) = 0$. Consequently, in a ring in which $3a = 0$ implies $a = 0$ and which satisfies the identity (1), the Jordan identity $(x^2, y, x) = 0$ must hold. Therefore a commutative, standard ring of characteristic not 3 is automatically a Jordan ring. It follows as an immediate Corollary to Theorem 1 that a simple, standard ring of characteristic not 3 is either a Jordan ring or associative. This is a generalization to rings of the theorem of Albert's mentioned in the introduction.

THEOREM 2. *If an accessible ring R has no nilpotent ideal other than zero, then it is isomorphic to a subdirect sum of an associative and a commutative ring.*

Proof. By assumption R can have no nilpotent ideal other than zero, so that $D = 0$. Hence R has no nilpotent elements in its center. Consider the natural homomorphism from R into $R/A \oplus R/B$. The kernel of this homomorphism is $A \cap B = 0$. Hence R is a subdirect sum of R/A and R/B. We have already noted that R/A is associative and that R/B is commutative. This completes the proof of the theorem.

The following is a direct consequence of Theorem 2. If an accessible ring R has a maximal nilpotent ideal I then R/I satisfies the conclusion of Theorem 2. Any expression involving elements of R, which would be automatically zero if the elements came from either an associative or a commutative ring, therefore must generate a nilpotent ideal of R. Of course the definition of accessibility requires only that two expressions, namely those occurring in (2) and (3), be zero.

The last result is concerned with a conventional type of decomposition, the introduction of a radical. Since the class of accessible rings includes the associative ones, the maximal nilpotent ideal will in general prove an unsatisfactory radical. We turn to a larger radical, namely the generalization of the Jacobson radical suggested by Brown **(2)**. From this paper it follows that an accessible ring is semi-simple if and only if it is isomorphic to a subdirect sum of primitive accessible rings. A ring is defined as primitive in case it possesses a regular maximal right ideal F, which contains no two-sided ideal of the ring other than the zero ideal.

We assert

THEOREM 3. *A semi-simple, accessible ring is a subidrect sum of primitive, accessible rings. A primitive, accessible ring is either commutative or associative.*

Proof. Only the second statement remains to be proved. Let R be a primitive, accessible ring and F a regular maximal right ideal of R which contains no two-sided ideal of R other than the zero ideal. The first step will be proving that R is prime. That is to say, if G and H are ideals of R such that $GH = 0$ and $G \neq 0$, then $H = 0$. We note that $G \not\subset F$, so that $R = F + G$. Then

$$RH = (F + G)H = FH + GH = FH \subset F.$$

For arbitrary elements x and y in R and h in H we have

$$(x, y, h) = xy \cdot h - x \cdot yh = xy \cdot h - xh' \in RH.$$

But $(x, y, h) = -(h, y, x)$, because of the flexible law, so that $(h, y, x) \in RH$. Finally, by means of (2), it can be shown that $(y, h, x) \in RH$. At this point it is easy to see that RH is an ideal of R. Since $RH \subset F$, then in fact $RH = 0$.

The regularity of F assures the existence of an element f in R with the property that for all x in R, $fx - x$ is always in F. Then in particular $fh - h = -h$ is an element of F. Consequently $H \subset F$. Since H is an ideal, $H = 0$. Since a prime ring has no nilpotent ideals other than the zero ideal it has no nilpotent elements in its center. As previously shown this implies that the ideals A, B of R have the property $AB = 0$. Hence either $A = 0$, in which case R is associative or $B = 0$, in which case R is commutative. This completes the proof.

References

1. A. A. Albert, *Power-associative rings*, Trans. Amer. Math. Soc., 64 (1948), 552–593.
2. Bailey Brown, *An extension of the Jacobson radical*, Proc. Amer. Math. Soc., 2 (1951), 114–117.

Ohio State University

ALGEBRAIC AND DIAGONABLE RINGS

M. P. DRAZIN

1. Introduction. In a well-known paper **(7)** Jacobson has shown how his structure theory for arbitrary rings can be applied to give more precise information about the so-called "algebraic" algebras. This specialization of his general theory is, however, perhaps not completely satisfying in that it deals only with algebras, i.e. rings admitting a *field* of operators, whereas neither the general structure theory nor the definition of the property of being "algebraic" seems to depend in any essential way on the precise nature of the operators.

In this paper we first show (§2) how, by suitably extending the algebraic concept to rings with arbitrary operators, Jacobson's theory of algebraic algebras can be carried over without difficulty to all "algebraic rings." Our definition of the algebraic property for arbitrary rings seems a natural one (and indeed almost inevitable if the link with π-regularity is to be preserved), and in §3 we establish some general results connected with this definitior. The first of these is unspectacular, and in any case applies only to algebras; it serves chiefly as a lemma for a theorem proved later (§5). However, the second result, whose hypothesis actually excludes fields as operators, is more surprising, having the corollary that *every ring algebraic over the integers and of zero characteristic must in fact be nil*; thus the algebraic property, as defined here with respect to arbitrary operator domains, can, for some choices of the operators, and in contrast with its more usual role of "weak finite-dimensionality," be a very strong one.

In the remaining sections we investigate various related questions. Thus in §4 we generalize the familiar result that a finite-dimensional matrix algebra over an algebraically closed field must be commutative whenever every matrix in the algebra can be reduced to diagonal form by a similarity transformation (allowed in the first instance to depend on the matrix); our generalization (which is applicable to algebras over *any* field, and indeed to arbitrary rings) has a certain topical interest in view of some recent work of Motzkin and Taussky. One of the new results in §5, while again referring only to algebras, generalizes Jacobson's result that every algebraic algebra without non-zero nilpotent elements, over a finite field, is necessarily commutative; we show in particular that the conclusion remains true even if non-zero nilpotent elements exist, provided these are all central. The earlier results of §5 are of a rather curious and apparently superficial type, but do nevertheless have some unexpected implications (e.g. that, if a π-regular, or in particular algebraic, ring R has all its nilpotent elements central, then the same is true of every homomorphic image of R).

Received October 4, 1955.

We recall Herstein's result **(5)** that *if, to each element x of a given ring R, there corresponds a polynomial $p_x(\lambda)$ with integral coefficients (and possibly a constant term) such that $x - x^2p_x(x)$ lies in the centre of R, then R must be commutative.* We shall refer to this as *Herstein's theorem*, and apply it in §5 and §6, where we show how certain analogous results, and a few special cases of a related conjecture of Herstein, can be deduced from our earlier work.

2. Preliminaries. Throughout, R will denote any associative ring, not necessarily commutative or containing a unit element, admitting an arbitrary commutative ring F of operators (i.e. endomorphisms α of the additive group of R, subject to $\alpha(xy) = (\alpha x)y = x(\alpha y)$ for all $x, y \in R$); we may suppose without loss of generality that F contains the identity operator. The case of a "ring without operators" is included in this scheme on taking F to be just the ring of integers (or an appropriate quotient ring). When we refer to subrings (etc.) of R these should always be understood as sub-F-rings (etc.), i.e. as being mapped into themselves by every operator in F.

If one seeks to introduce an analogue, at this level of generality, of the property of an algebra of being "algebraic over its field of operators," one may (cf. **3**) think first of calling an element x of R algebraic over F if a positive integer n and elements $\alpha_1, \ldots, \alpha_n$ of F, not all zero, exist satisfying

$$\alpha_1 x + \alpha_2 x^2 + \ldots + \alpha_n x^n = 0. \tag{1}$$

This of course reduces to Jacobson's definition when F is a field. However, this form is unsatisfactory from many points of view, as will become clearer below; we note for the present that it would not even enable us to carry over to rings the well-known property of algebraic algebras of having nil Jacobson radical. We therefore adopt a more stringent defining condition: we shall now call x *algebraic* (*over the ring F of operators*) whenever $\alpha_1, \ldots, \alpha_n$ exist as above but with the further property that the first non-vanishing α_i is the identity operator, i.e. only if x satisfies an equation of the ("lower monic") form

$$x^m + \alpha_{m+1}x^{m+1} + \ldots + \alpha_n x^n = 0; \tag{2}$$

if R happens to be an algebra, i.e. if F is a field, then this can of course always be arranged (on multiplying through by the inverse of the lowest non-zero coefficient) whenever x satisfies the formally weaker condition (1). Another equivalent form of our new definition is the following: x is algebraic if we can find a positive integer $m = m(x)$, and an element $a = a(x)$ of the subring generated by x, such that $x^m = x^m a$. We call R itself algebraic over F if each $x \in R$ is algebraic over F.

It is a straightforward matter to check that all the principal arguments and results of Jacobson's paper **(7)** on algebraic algebras are valid, with only slight verbal changes, for the wider class of algebraic rings; we omit the details.

It is important to bring out into the open a point which might otherwise give rise to misunderstandings later. Given a ring R over F, we may regard

any homomorphic image $R^* = R/T$ of R as again a ring over F by defining $\alpha x^* = (\alpha x)^*$ in the usual way. However, if we agree to regard two operators as equal relative to a given ring (which admits them both) if and only if they have the same effects on each element of the ring, then the operator set on R^* is, strictly, not F but the factor ring $F^* = F/G$, where G denotes the ideal of F consisting of all $\alpha \in F$ such that $\alpha R \leqslant T$. This distinction, vacuous when F is a field, can nevertheless be vital for more general operator rings F (particularly when their cardinals or characteristics are in question). Also, if we had chosen to define algebraic elements by means of (1), we could not have asserted that R being algebraic over F implies that R^* is algebraic over F^* (since some $x \in R$ might satisfy only equations (1) in which each coefficient $\alpha_i \in G$); however, using (2) ensures homomorphism-invariance for the algebraic property (since the identity element of F maps onto that of F^*).

In view of these remarks, it is not strictly true to say that every ring may be regarded as a ring over the ring I of integers: in fact this will be legitimate for a given ring R if and only if, for each positive integer k, an element x exists in R such that $kx \neq 0$. However, it is convenient and in practice not seriously confusing to be a little inexact in this connexion: we shall allow ourselves the customary liberty of regarding any ring R as a ring over I (rather than some quotient ring of I). Thus, for example, any algebra algebraic over a finite field of prime order will be regarded also as algebraic over the integers.

Our definition of the algebraic property via (2), while fulfilling most reasonable requirements, does have the slight technical disadvantage of carrying with it no immediately available concept of a minimal polynomial; for, among the polynomials satisfying (2), there is in general more than one of minimal "lower degree" m (even if we demand that $n - m$ be also minimal). However, at least when F is an integral domain, we can get something with most of the usual properties by returning to (1).

Let R be a ring with arbitrary operators F, and x any element of R algebraic over F. Then x satisfies an equation of the form (2), and *a fortiori* satisfies equations of the form (1), i.e. there are non-zero polynomials $f(\lambda)$ over F, without constant terms, such that $f(x) = 0$. Among such polynomials $f(\lambda)$, all those of minimal degree (there will in general be several, possibly infinitely many) will be called ***minimal polynomials for x over F***. We note two relevant lemmas; the first is standard and leads immediately to the second.

LEMMA 2.1. *Let $f(\lambda)$, $g(\lambda)$ be arbitrary formal polynomials over a given commutative ring F, with leading terms $\alpha\lambda^n$, $\beta\lambda^k$ respectively. Then there are polynomials $q(\lambda)$, $r(\lambda)$ over F, with $r(\lambda)$ zero or of degree strictly less than n, such that*

$$\alpha^k g(\lambda) = q(\lambda) f(\lambda) + r(\lambda).$$

LEMMA 2.2. *Let x be any algebraic element of a given ring R over F, and let $f(\lambda)$ be any minimal polynomial for x over F, say with leading term $\alpha\lambda^n$. Then,*

given any polynomial $g(\lambda)$ *over* F, *of degree* k *say, such that* $g(x)=0$, *there is a polynomial* $q(\lambda)$ *over* F *(possibly with constant term) such that*

$$\alpha^k g(\lambda) = q(\lambda) f(\lambda).$$

Of course Lemma 2.2 is of value only when we can be sure that $\alpha^k \neq 0$. If F is an integral domain we even have some measure of uniqueness ("up to scalar factors") for our minimal polynomials: for, if f, g are two such, with leading terms $\alpha\lambda^n$, $\beta\lambda^k$, then, by Lemma 2.2, polynomials $p(\lambda)$, $q(\lambda)$ exist such that $\alpha^k g = qf$, $\beta^n f = pg$, whence $\alpha^k\beta^n f = pqf$. Thus, for an integral domain F, since f is not identically zero, we have $\alpha^k\beta^n = p(\lambda)\,q(\lambda)$, and so $p(\lambda)$, $q(\lambda)$ must both be non-zero constants; in other words, any two minimal polynomials of a given element x algebraic over an integral domain must have a common non-zero scalar multiple.

3. Some general properties of algebraic rings. Our first theorem (which will find a use later) is a direct adaptation of a result from elementary algebraic number theory:

THEOREM 3.1. *Let* F *be a field, algebraic over a given subfield* F_0. *Then every algebra algebraic over* F *is algebraic over* F_0.

Proof. Let R be any ring over F, and x any element of R algebraic over F, say satisfying (2) above, with each $\alpha_i \in F$. Since F is a field, we can single out from the non-zero α_i that one, say α_q, with greatest index q, multiply through by α_q^{-1}, and write

$$x^q = \beta_1 x + \ldots + \beta_{q-1}x^{q-1},$$

where $q > m > 1$, each $\beta_i \in F$. Hence, if we denote the field $F_0(\beta_1, \ldots, \beta_{q-1})$ by K, then the algebra $K[x]$ is finite-dimensional over K, while also K is itself a finite extension of F_0 (since each β_i is algebraic over F_0). Thus $K[x]$ can be regarded as a finite-dimensional algebra over F_0 (its dimension as such being given by $\dim(K[x]\colon K)\dim(K\colon F_0)$), that is, x is algebraic over F_0, as required.

We come now to some of our principal results.

THEOREM 3.2. *Let* F *be any (commutative) integral domain, not a field but having a unit element, and let* R *be any ring algebraic over* F. *Then, given any element* x *of* R, *any equation of the form*

$$\alpha_m x^m + \alpha_{m+1}x^{m+1} + \ldots + \alpha_n x^n = 0 \tag{3}$$

with each $\alpha_i \in F$ *and* $\alpha_m \neq 0$ *(and of course some such relation holds) implies the existence of a non-zero element* α *of* F *such that* $\alpha x^m = 0$.

Proof. We can write $\alpha_m x^m = x^m a$, where a is in the subring of R generated by x; then $\alpha_m{}^j x^m = x^m a^j$ $(j = 1, 2, \ldots)$, and so, taking $j = m+1$, we can find $b \in R$ such that $\alpha_m{}^{m+1}x^m = x^m b x^m$. Defining $e = x^m b$, we then have

$$\alpha_m^{m+1}x^m = ex^m, \quad \alpha_m^{m+1}e = e^2.$$

Now, for any $\beta \in F$, since R is algebraic, we can find a positive integer t_β and a polynomial $k_\beta(\lambda)$ over F such that $(\beta e)^{t_\beta} = (\beta e)^{t_\beta+1}k_\beta(\beta e)$. Also, by use of the relation $e^2 = \alpha_m{}^{m+1}e$, we can express $e^2k_\beta(\beta e) = \theta_\beta e$ for some $\theta_\beta \in F$, so that $\beta^{t_\beta}e^{t_\beta} = \beta^{t_\beta+1}\,\theta_\beta e^{t_\beta}$, that is

$$0 = \beta^{t_\beta}(1 - \beta\theta_\beta)\,e^{t_\beta} = \beta^{t_\beta}(1 - \beta\theta_\beta)\,\alpha_m^{(m+1)(t_\beta-1)}e;$$

consequently, for each $\beta \in F$,

$$\beta^{t_\beta}(1 - \beta\theta_\beta)\,\alpha_m^{(m+1)t_\beta}x^m = 0.$$

Thus either $\alpha \equiv \beta^{t_\beta}(1 - \beta\theta_\beta)\alpha_m{}^{(m+1)t_\beta}$ is non-zero for some $\beta \in F$, as required, or else $\beta^{t_\beta}(1 - \beta\theta_\beta)\alpha_m{}^{(m+1)t_\beta} = 0$ for all $\beta \in F$; and in this latter case (since $\alpha_m \neq 0$ and F is an integral domain) we should have $1 = \beta\theta_\beta$ for each non-zero $\beta \in F$, contrary to our hypothesis that F is not a field.

COROLLARY 3.1. *Let R be any ring of characteristic zero. Then R is algebraic over the ring of integers if and only if R is nil.*

Proof. The ring of integers satisfies the conditions on F in Theorem 3.2, so, if R is algebraic over the integers, then, to each $x \in R$, there correspond a non-zero integer $\alpha = \alpha(x)$ and a positive integer $m = m(x)$ such that $\alpha x^m = 0$; and, since R has characteristic zero, this implies that $x^m = 0$, whence R is nil. The converse is obvious.

Theorem 3.2 may be regarded as generalizing the known fact (**7**, Theorem 11) that, if every element of a ring R satisfies $x^{n(x)} = x$ for some integer $n(x) > 2$, then every element of R has finite additive order; indeed, for any element x of a ring R satisfying this more stringent condition, and any admissible operator ring F, our argument shows that either an element α of F exists such that $\alpha x = 0$, $\alpha R \neq 0$, or F has the same property as R (so that, if F is an integral domain, it must be an algebraic field of prime characteristic).

The argument of Theorem 3.2 can easily be modified to show that, with F, R as before, every regular element of R has a non-zero annihilator in F. We can also, without appreciably more trouble, prove the following generalization of Theorem 3.2 (cf. **11**):

THEOREM 3.3. *Let F be any integral domain, R any ring over F, and x any element of R. Suppose also that there exists a non-zero element $\pi = \pi(x)$ of F such that, to each element y of the subring of R generated by x, corresponds a non-zero polynomial $g_y(\lambda)$ over F, whose lowest non-zero coefficient is not divisible by π, such that $g_y(y) = 0$. Then any equation of the form* (3) *with each $\alpha_i \in F$ and $\alpha_m \neq 0$ implies that a non-zero element α of F exists satisfying $\alpha x^m = 0$.*

Proof. As in the proof of Theorem 3.2, we can find an element e of the subring generated by x such that $\alpha_m{}^{m+1}x^m = ex^m$, $\alpha_m{}^{m+1}e = e^2$. By our hypothesis, for each $\beta \in F$, there is a polynomial over F of the form $g_{\beta e}(\lambda) = \gamma_\beta\lambda^{t_\beta} - \lambda^{t_\beta+1}k_\beta(\lambda)$, with γ_β not divisible by π, such that $g_{\beta e}(\beta e) = 0$. As before, we deduce that, for each $\beta \in F$, an element θ_β of F exists such that

$$\beta^{t_\beta}(\gamma_\beta - \beta\theta_\beta)\,\alpha_m^{(m+1)t_\beta}x^m = 0.$$

Finally, taking $\beta = \pi$, since $\pi \neq 0$, $\alpha_m \neq 0$ and since γ_r is not divisible by π, we can be sure that $\alpha \equiv \pi^{t_r}(\gamma_r - \pi\theta_r)\alpha_m^{(m+1)t_r} \neq 0$.

It is hardly necessary to mention that the existence of an element π of F satisfying the conditions of Theorem 3.3 ensures that F cannot be a field. As a corollary of Theorem 3.2 itself (or more generally of Theorem 3.3) it is obvious that any minimal polynomial of x must have the monomial form $\alpha\lambda^n$. This is not difficult to see even under a hypothesis substantially weaker than that all elements of the *subring* generated by x be algebraic, as we show next:

THEOREM 3.4. *Let F be any integral domain, not a field but having a unit element, let R be any ring over F, and x a given element of R. Then, if every F-multiple γx of x is algebraic over F, and if*

$$h(\lambda) = \alpha_m\lambda^m + \alpha_{m+1}\lambda^{m+1} + \ldots + \alpha_n\lambda^n,$$

with $\alpha_m \neq 0$, $\alpha_n \neq 0$, is a given minimal polynomial for x over F, we must have $m = n$ (so that $\alpha_m x^m = 0$).

Proof. Suppose by way of contradiction that $m \neq n$, that is, $n - m \geqslant 1$, and let β be an arbitrary non-zero element of F (fixed throughout the ensuing argument). Then

$$0 = \beta^n\alpha_n^{n-1}h(x) = \beta^{n-m}\alpha_n^{n-m-1}\alpha_m(\alpha_n\beta x)^m + \beta^{n-m-1}\alpha_n^{n-m-2}\alpha_{m+1}(\alpha_n\beta x)^{m+1} + \ldots + \beta\alpha_{n-1}(\alpha_n\beta x)^{n-1} + (\alpha_n\beta x)^n,$$

and so, defining $y = \alpha_n\beta x$ and

$$f(\lambda) = \beta^{n-m}\alpha_n^{n-m-1}\alpha_m\lambda^m + \beta^{n-m-1}\alpha_n^{n-m-2}\alpha_{m+1}\lambda^{m+1} + \ldots + \beta\alpha_{n-1}\lambda^{n-1} + \lambda^n,$$

we have $f(y) = 0$. Indeed, $f(\lambda)$ is a *minimal* polynomial for y (since, F being an integral domain and $\alpha_n\beta$ being non-zero, if y satisfied an equation of lower degree, so would x).

Now, y being an F-multiple of x, our hypothesis assures us of the existence of a positive integer t and a polynomial $k(\lambda)$ over F such that $y^t = y^{t+1}k(y)$. Thus, by Lemma 2.2 (with $\alpha = 1$), there is a polynomial $q(\lambda)$ over F such that

$$\lambda^t - \lambda^{t+1}k(\lambda) = f(\lambda)\, q(\lambda)$$

identically. Since $\alpha_m \neq 0$, comparison of coefficients of λ^t on either side gives $1 = \beta^{n-m}\alpha_n^{n-m-1}\alpha_m\xi$, where $\xi = \xi_\beta$ is the lowest non-zero coefficient of $q(\lambda)$, that is $1 = \beta\theta_\beta$, where

$$\theta_\beta = \beta^{n-m-1}\alpha_n^{n-m-1}\alpha_m\xi_\beta.$$

But, since β was an arbitrary non-zero element of F, this would contradict our hypothesis that F is not a field; thus in fact $m = n$, as required.

There is naturally an extension of Theorem 3.4 along the lines of Theorem 3.3, but we shall not state it formally. However, we note the (trivial and known) corollary that, if z is a complex number such that z/β is an algebraic

integer (in the usual number-theoretic sense) for every positive integer β, then $z = 0$; to see this, one has only to suppose the contrary, and take $x = 1/z$ in Theorem 3.4.

4. Diagonable rings. On being given any positive integer q and on writing 1_q for the unit $q \times q$ matrix, it is customary to call a $q \times q$ matrix x, with elements in a given field F, *diagonable over* F if distinct elements $\beta_1, \ldots, \beta_s$ of F exist such that

$$(x - \beta_1 1_q) \ldots (x - \beta_s 1_q) = 0$$

(where s can be any positive integer). There are several well-known alternative forms for this definition (e.g. in terms of the existence of a non-singular $q \times q$ matrix b over F such that $b^{-1}xb$ is diagonal). We shall adopt the following (obviously equivalent) form: x is diagonable over F if and only if there are distinct elements $\gamma_1, \ldots, \gamma_t$ of F such that

$$x(\gamma_1 x + 1_q) \ldots (\gamma_t x + 1_q) = 0. \tag{4}$$

It will be noted that we have not required F to be algebraically closed; indeed, our definition remains significant for *any* (*commutative*) *ring* F. Further, since the unit matrix 1_q now occurs only in a purely formal way (i.e. can be got rid of by multiplying out the factors in (4)), we may apply the definition to *any* ring R admitting the operators F (i.e. not merely to rings of square matrices over F). If every element of a ring R over F is diagonable over F, we shall say that R is itself diagonable over F. Obviously every diagonable ring over F is algebraic over F.

Motzkin and Taussky **(10)** showed that, if x, y are given $q \times q$ matrices over an algebraically closed field F, and if also $\alpha x + \beta y$ is diagonable over F for all choices of α, β in F, then $xy = yx$ (whence it is easy to deduce the existence of a non-singular $q \times q$ matrix b over F reducing x and y simultaneously to diagonal forms $b^{-1}xb$, $b^{-1}yb$). Their proof (a geometrical one) is long; and, since hypotheses are made only about the F-module generated by x and y, ring-theoretic methods are perhaps not very suitable for dealing with the problem. However, if we are prepared to extend the diagonability hypothesis to all "non-commutative polynomials" in x and y, then the proof that x, y commute becomes almost trivial; indeed, for rings with arbitrary operators, we shall show in our next theorem that diagonability always implies commutativity. The proof depends on a familiar property of strongly regular rings; for completeness, we first derive this property, and indeed something more general, in the following lemma (which will in any case be needed later on in §6):

LEMMA 4.1. *Let R be any ring in which, to each pair of elements x, y, there corresponds a non-negative integer r such that xy^r is in the right ideal of R generated* (*over the given operator ring F*) *by y and x^2. Then, if J denotes the Jacobson radical of R, R/J is a subdirect sum of division rings.*

Proof. We know from Jacobson's structure theory that R/J is a subdirect sum of primitive rings, each of which is a homomorphic image of R/J and hence of R; and each of these primitive rings inherits the (clearly homomorphism-transitive) hypothesis on R. Thus it will be enough to show that if R is itself primitive then R must be a division ring.

To call R primitive is the same as to say that R is isomorphic with a dense ring M of linear transformations of a vector space V over a division ring D. We shall denote the result of operating on $v \in V$ with $x \in M$ by vx (i.e. regard M and D as operating on V from the right), and have only to show that V cannot contain two elements v_1, v_2 independent with respect to D. But, in the contrary case, since M is dense, we could choose x, y in M so that

$$v_1x = v_2,\ v_2x = 0,\ v_1y = 0,\ v_2y = v_2;$$

then, for any $\alpha, \beta \in F$, any $a, b \in R$, and any non-negative integer r,

$$v_1(xy^r - \alpha x^2 - x^2a - \beta y - yb) = v_1x(y^r - \alpha x - xa) - 0 = v_2y^r = v_2 \neq 0$$

(by the D-independence of v_1, v_2). But our hypothesis on R asserts that, x, y being chosen, we can find α, β, a, b, r such that $xy^r - \alpha x^2 - x^2a - \beta y - yb = 0$; thus we have our desired contradiction.

THEOREM 4.1. *Every diagonable ring is commutative.*

Proof. Given any element x of a diagonable ring R, then, on taking $\gamma_1, \ldots, \gamma_t$ as in equation (4) above and on writing

$$(\gamma_1\lambda + 1) \ldots (\gamma_t\lambda + 1) = 1 - \lambda g(\lambda),$$

$g(\lambda)$ is a polynomial over F (possibly with constant term), and $x = x^2h(x)$, where $h(\lambda) = \lambda g^2(\lambda)$ is a polynomial over F without constant term (so that $h(x)$ is well-defined). Thus, given any $x \in R$, we can find an element $a = h(x)$ of R such that $x = x^2a$. In other words, every diagonable ring R is strongly regular and hence semi-simple in Jacobson's sense, and so, by Lemma 4.1 (with $r = 0$), R is a subdirect sum of division rings, each of which is a homomorphic image of R and consequently diagonable. But a diagonable division ring is obviously commutative, so we deduce that R must in fact be a subdirect sum of fields.

5. Additive functions on π-regular rings. We recall (cf. **8**) that an element x of a ring R is said to be *π-regular in* R if a positive integer $s = s(x)$ and an element $b = b(x)$ of R exist satisfying $x^s = x^sbx^s$. Given any elements x, y of a ring, we shall use $[x, y]$ to denote their additive commutator $xy - yx$.

THEOREM 5.1. *Let R be any ring, let $\mathfrak{S}$ be any given set with a transitive binary relation $\prec$ defined on it, and let $\mathcal{M}$ be any set of mappings of R into $\mathfrak{S}$. Then, if we denote by $M(x)$ the result of operating on a typical element x of R by a typical element M of $\mathcal{M}$, the statement* (i) *to each choice of x in R and M in*

$\mathscr{M}$ there correspond $c \in R$, $N \in \mathscr{M}$ and an integer $t > 2$ such that $[x^t, c] = 0$ and $M(x) \prec N(x^t c)$, implies (ii) *$M(z) \prec M(0)$ for every $M \in \mathscr{M}$ and every nilpotent element z of R.*

Conversely, if (ii) *holds, then* (iii) *for any given π-regular element x of R, say with $x^s = x^s b x^s$, we have $M(x - x^{s+1}b) \prec M(0)$ for all $M \in \mathscr{M}$.*

Proof. Suppose first that (i) holds. Then, given any $x = x_0$ in R and any $M = M_0$ in $\mathscr{M}$, we can find a sequence of integers $t_j > 2$, a sequence c_j of elements of R and a sequence of mappings $M_j \in \mathscr{M}$ ($j = 1, 2, \ldots$) such that

$$x_j = x_{j-1}^{t_j} c_j,\ [x_{j-1}^{t_j}, c_j] = 0,\ M_{j-1}(x_{j-1}) \prec M_j(x_j) \qquad (j = 1, 2, \ldots).$$

Since $\prec$ is transitive on $\mathscr{M}$, $M(x) = M_0(x_0) \prec M_j(x_j)$ ($j = 1, 2, \ldots$), while a simple induction argument shows that

$$x_j = x^{t_1 \ldots t_j} c_1^{t_2 \ldots t_j} c_2^{t_3 \ldots t_j} \ldots c_{j-1}^{t_j} c_j \qquad (j = 1, 2, \ldots);$$

combining these two remarks we obtain (ii) on taking j sufficiently large.

To prove that (ii) implies (iii) we notice that $(x - x^{s+1}b)^s$ can be written in the form $x^s + x^s d$ (for a suitably chosen $d \in R$), so that

$$(x - x^{s+1}b)^{s+1} = (x - x^{s+1}b)\, x^s + (x - x^{s+1}b)\, x^s \,.\, d;$$

also, if $x^s = x^s b x^s$, then

$$(x - x^{s+1}b)\, x^s = x^{s+1} - x^{s+1} b x^s = x^{s+1} - x \,.\, x^s = 0,$$

so that $x - x^{s+1}b$ is nilpotent, and (ii) gives $M(x - x^{s+1}b) \prec M(0)$.

It is perhaps worth remarking that, if $x^s = x^s b x^s$ and we write $s = 2r - 1 + \delta$, where r is a positive integer and $\delta = 0$ or 1, then one can show (only slightly less easily than in the second part of the proof above) that $(x - x^{r+\delta} b x^r)^s = 0$, so that (ii) also implies $M(x - x^{r+\delta} b x^r) \prec M(0)$; however, this fact seems to be less useful in applications.

We have set out Theorem 5.1 in the very general (and accordingly rather bogus-looking) form above in order to highlight the essential argument, which will be successively more and more obscured in our next theorems (where we return to earth, and make the "converse" more worthy of the name, by specializing $\mathfrak{S}$, $\mathscr{M}$). We shall mean by an *additive function on R* any mapping, say $f: x \to f(x)$, of R into itself such that $f(x + y) = f(x) + f(y)$ for all $x, y \in R$; in particular, $f(0) = 0$. We do not require that $f(\alpha x) = \alpha f(x)$ for admissible operators α.

THEOREM 5.2. *Let R be any ring, and $\mathscr{L}$ any set of additive functions on R. Then the statement* (i) *to each choice of x in R and f in $\mathscr{L}$ there correspond $c \in R$, $g \in \mathscr{L}$ and an integer $t > 2$ such that $[x^t, c] = 0$ and such that $f(x)$ is in the two-sided ideal of R generated by $g(x^t c)$, implies* (ii) *$f(z) = 0$ for every $f \in \mathscr{L}$ and every nilpotent element z of R.*

Conversely, if (ii) *holds, then* (iii) *for any given π-regular element x of R, say with $x^s = x^s b x^s$, we have $f(x) = f(x^{s+1}b)$ for every $f \in \mathscr{L}$.*

Proof. For any $x \in R$, $f \in \mathscr{L}$, let $M_f(x)$ denote the two-sided ideal of R generated by the element $f(x)$ of R. Then Theorem 5.2 is just the special case of Theorem 5.1 with $\mathfrak{S}$ chosen as the set of all two-sided ideals of R, ordered in the natural way by inclusion, and with $\mathscr{M}$ chosen as the set of all mappings $M_f: x \to M_f(x)$.

We recall that an element x of a ring R is said to be *strongly regular in* R if an element $a = a(x)$ of R exists such that $x = x^2a$.

THEOREM 5.3. *Every π-regular ring without non-zero nilpotent elements is strongly regular.*

Proof. This follows at once from the second part of Theorem 5.2 on taking $\mathscr{L}$ to consist of the single function $f: x \to f(x) = x$. Alternatively and more directly, going back to the proof of Theorem 5.1, we have merely to observe that $x^s = x^s bx^s$ implies $(x - x^{s+1}b)^{s+1} = 0$, so that, if R has no non-zero nilpotent elements, then $x - x^{s+1}b = 0$, that is

$$x = x^2(x^{s-1}b).$$

Conversely, if R is strongly regular, then (independently of the π-regularity hypothesis) of course R can obtain no non-zero nilpotent element. Thus we see that, *among π-regular rings, the property of having no non-zero nilpotent elements is homomorphism-invariant.*

From now on all we shall need of what has already been proved in this section is the following consequence of Theorem 5.2:

THEOREM 5.4. *For any ring R, the statement* (i) *to each choice of x, y in R there correspond $c \in R$ and an integer $t \geqslant 2$ such that* $[x^t, c] = [x - x^tc, y] = 0$, *implies* (ii) *every nilpotent element of R is central.*

Conversely, if (ii) *holds, and x is any given π-regular element of R, say with $x^s = x^sbx^s$, then* (i) *holds, for this x and all y, with $c = b$ and $t = 2$ if $s = 1$, and with $c = xb$ and $t = s$ otherwise.*

Proof. The first part is essentially the special case of the corresponding part of Theorem 5.2 with $\mathscr{L}$ chosen as the set of "commutator functions" $f_y: x \to [x, y]$ (one such function being associated with each $y \in R$); indeed, we have thrown away some generality elsewhere by writing $[x - x^tc, y] = 0$ in (i) rather than the weaker "for some z in R, $[x, y]$ lies in the two-sided ideal of R generated by $[x^tc, z]$."

To prove the converse, we quote from **(2)** that (ii) implies that every idempotent element e of R is central. Taking $e = x^sb$, we deduce that $x^s = x^{2s}b$; in a similar way, we see that $x^s = bx^{2s}$. Hence

$$x^sb = bx^{2s} \,.\, b = b \,.\, x^{2s}b = bx^s;$$

also, by the converse part of Theorem 5.2, $x - x^{s+1}b$ is central, so the proof is complete.

In particular, we have proved that (i) and (ii) are equivalent in any π-regular ring. For the special case of rings algebraic over the integers, the two parts of Theorem 5.4 are implicit in Herstein's papers (**5**, Lemma 3; **6**) respectively. We note also the following immediate consequence of Theorem 5.4:

COROLLARY 5.1. *Among π-regular rings, the property of having all nilpotent elements central is preserved under homomorphism (even under homomorphisms which do not commute with the given operators).*

This corollary cannot of course be extended to arbitrary rings (consider for example the free ring R generated over the integers by two non-commutative indeterminates, and the natural homomorphism of R onto $R/(4R)$).

Combining Herstein's theorem (quoted in the Introduction) with the converse part of Theorem 5.4, we have (since every algebraic ring is clearly π-regular)

THEOREM 5.5. *Let R be a given ring algebraic over the integers (or any quotient ring), and suppose that every nilpotent element of R is central. Then R is commutative.*

This was previously pointed out by Herstein (**5**), and generalizes a result of Arens and Kaplansky (**1**, Theorem 4.2). They proved commutativity for any ring R, necessarily without non-zero nilpotent elements, in which each element x has finite non-zero additive order and satisfies an equation of the form

$$\alpha_1 x + \ldots + \alpha_n x^n = 0, \tag{1}$$

with $\alpha_1, \ldots, \alpha_n$ integral and $\alpha_n x^n \neq 0$. For in these circumstances every element has squarefree characteristic, so that R is the restricted direct sum of $R_{(p)}$, where p takes all prime values and $R_{(p)}$ denotes the set of all $x \in R$ with $px = 0$; and it is easy to see that each $R_{(p)}$ is algebraic over the integers (in our sense) and without non-zero nilpotent elements. Restrictions on the additive orders of elements of R are no longer in evidence in the statement of Theorem 5.5; however, Theorem 3.2 shows that this aspect of generalization of the result of Arens and Kaplansky is illusory.

We should naturally like to have something similar to Theorem 5.5 valid for rings with more general operators than the integers. Such a generalization would of course follow for a given operator ring F if Herstein's theorem could be extended to allow elements of F as coefficients in $p_x(\lambda)$. Consideration of the quaternion algebra over the reals sets a limit on such hopes, but, by Theorems 3.1 and 5.5, we do have at least the following generalization of Jacobson's result (**7**, Theorem 9) mentioned in the Introduction:

THEOREM 5.6. *Let F be any field of non-zero characteristic algebraic over its prime subfield (in particular, any finite field). Then, if a given algebra R algebraic over F has all its nilpotent elements central, R is commutative.*

It will be noted that the hypotheses on F imply that F is a perfect field; however, the quaternions show that the result does not hold for all perfect fields F.

We note also the following analogous, and more elementary, result:

THEOREM 5.7. *Let F be any algebraically closed field. Then, if a given algebra R algebraic over F has all its nilpotent elements central, R is commutative.*

Proof. Given any element x of R, then, since R is algebraic, x generates a finite-dimensional subalgebra over F, and, since F is algebraically closed, consequently, by the theory of the classical canonical form, we can write

$$x = f + \sum_i \alpha_i e_i,$$

where f is a nilpotent (and hence central) element of R, the e_i are idempotent elements of R, and the α_i are in F. Thus, to prove R commutative, it would be enough to show that all idempotent elements of R commute with one another. But in fact, by **(2)** again, the hypothesis that all nilpotent elements are central implies (in any ring) that every idempotent element is central, so the result follows.

6. H-rings. We now turn to some questions arising from Herstein's theorem. Herstein's method of proof was to settle first the division ring case (which he succeeded in doing by a comparatively short argument), and then to show (by a rather lengthy sequence of lemmas) how the result for arbitrary rings can be reduced to this special case. Herstein has conjectured (in a letter to the writer) that if, to each element x of a given ring R, there corresponds an element a of R such that $x - x^2a$ is central, then R is a subdirect sum of a commutative ring and a (possibly vacuous) set of division rings; we shall refer to this as *Herstein's conjecture.*

This conjecture can reasonably be thought of as generalizing Herstein's theorem, since any division ring D occurring as a subdirect summand of R is necessarily a homomorphic image of R, so that, if a is always a polynomial in the $x \in R$ to which it corresponds, then a similar statement holds for D (while, as we have noted, the division ring case of Herstein's theorem takes up only a small part of the proof). Further, the conjecture, if true, would have over the theorem the advantage that its (much weaker) hypothesis does not involve any restriction on the operators, whereas the quaternion ring shows that the theorem as originally stated definitely does not extend to rings with arbitrary operators (rather than the integers). Thus the conjecture embodies as much as one could hope to be true in the general case and also, essentially, in the case of integer operators first considered by Herstein; if the conjecture could be substantiated, the theorem (and most of its subsequent ramifications) could be deduced from it in a comparatively trivial way.

We shall in fact consider here only the case in which $x^2a = x^tc = cx^t$, where $c \in R$ and t is some integer with $t \geqslant 2$, but we can afford to weaken the centrality condition slightly. Formally, we call a given ring R an *H-ring* if, to each pair x, y in R, there correspond $c = c(x, y) \in R$ and an integer $t = t(x, y) \geqslant 2$ such that

$$[x^t, c] = [x - x^tc, y] = 0.$$

Certain of Herstein's arguments can be straightforwardly generalized to apply to these rings; since every division ring is an H-ring (e.g. with $c = x^{-1}$ for $x \neq 0$ and otherwise arbitrary) we cannot hope to prove all H-rings commutative, and we shall be chiefly concerned with side-conditions sufficient to ensure commutativity (cf. Theorem 4.1 above). Our next result shows that all H-rings have a certain property which would follow as an immediate consequence of the truth of Herstein's conjecture; and, conversely, that, when we restrict attention to π-regular rings, this property actually characterizes the H-rings:

THEOREM 6.1. *Every H-ring has all its nilpotent elements central. Conversely, if a given ring R is π-regular and all its nilpotent elements are central, then R is an H-ring and c, t can be chosen independently of y; also, if R is algebraic (over some given ring F of operators), then, corresponding to each $x \in R$, there is a polynomial $p_x(\lambda)$ over F such that $x - x^2p_x(x)$ is central.*

This is, essentially, just a partial restatement of a special case of Theorem 5.4 in H-ring terminology.

Extending slightly concepts which have been used by Goldhaber and Whaples **(4)** and by McLaughlin and Rosenberg **(9)**, we shall say that a commutative ring F is *quasi-algebraically closed* if every division ring algebraic over F (or over a factor ring of F) is commutative. Obviously every algebraically closed field is quasi-algebraically closed; and, by Theorems 5.5 and 5.6, the property of being quasi-algebraically closed is also shared by the ring of integers (with all its quotient rings), and by every finite field.

If, in Lemma 4.1, F is quasi-algebraically closed and R is algebraic over F, then clearly R/J is a subdirect sum of fields, and so (since J is nil in any π-regular ring) R is commutator-nil, i.e. the two-sided ideal of R generated by all the commutators $[x, y]$ with $x, y \in R$ is a nil ideal;[1] and clearly every H-ring satisfies the hypothesis of Lemma 4.1 (with $r = 1$). If Herstein's conjecture were true, we should even have commutativity for all algebraic H-rings over quasi-algebraically closed operator rings F. Not every algebraic H-ring is commutative (consider again the quaternions), but, by combining Theorem 6.1 with Herstein's theorem, and also with Theorems 3.1 and 5.7, we find

[1]Clearly this conclusion still holds good even if R is given as only π-regular (rather than actually algebraic) provided that *every* division ring over F is commutative; a variety of analogous results can be obtained by weakening the hypothesis on either R or F and correspondingly strengthening the hypothesis on the other.

THEOREM 6.2. *Every H-ring algebraic over the integers, or over any finite or algebraically closed field, is commutative.*

More generally, if Herstein's theorem could be extended to allow given operators F as coefficients in $p_x(\lambda)$, then we could similarly show that every H-ring algebraic over this particular F is commutative. Commutativity (and hence local finiteness) would then of course follow for every division ring algebraic over F; and this is the same as to say that F is quasi-algebraically closed. Thus Herstein's theorem definitely cannot be extended to any non-quasi-algebraically closed F.

Without prejudging how far Herstein's theorem does extend, or whether his conjecture is in fact true, we can at least show that every H-ring algebraic over a quasi-algebraically closed *field* F must be locally finite. For the algebraic condition on R makes J nil and consequently (by Theorem 6.1) central, while we have previously seen (from Lemma 4.1) that R/J must be commutative. Then R/J and J, being commutative algebraic algebras over F, are both locally finite over F, whence, by (**7**, Theorem 15), R is itself locally finite, as we asserted.

Now commutativity for R is equivalent to that of all its doubly generated subrings; also these are finite-dimensional over F by what we have just proved, and are H-rings in view of the last part of Theorem 6.1. Thus, to prove commutativity for all H-rings algebraic over a given quasi-algebraically closed field F, it is enough to do so only for finite-dimensional R; this is easy when F is also perfect (and, more generally, whenever R can be expressed as a supplementary sum $R = S + J$ with $S \cong R/J$).

REFERENCES

1. R. Arens and I. Kaplansky, *Topological representation of algebras*, Trans. Amer. Math. Soc. *63* (1948), 457–481.

2. M. P. Drazin, *Rings with central idempotent or nilpotent elements* (to appear).

3. M. P. Drazin and K. W. Gruenberg, *Commutators in associative rings*, Proc. Cambridge Phil. Soc., *49* (1953), 590–594.

4. J. K. Goldhaber and G. Whaples, *On some matrix theorems of Frobenius and McCoy*, Can. J. Math., *5* (1953), 332–335.

5. I. N. Herstein, *The structure of a certain class of rings*, Amer. J. Math., *75* (1953), 864–871.

6. ———, *A note on rings with central nilpotent elements*, Proc. Amer. Math. Soc., *5* (1954), 620.

7. N. Jacobson, *Structure theory for algebraic algebras of bounded degree*, Ann. Math., *46* (1945), 695–707.

8. N. H. McCoy, *Generalized regular rings*, Bull. Amer. Math. Soc., *45* (1939), 175–178.

9. J. E. McLaughlin and A. Rosenberg, *Zero divisors and commutativity of rings*, Proc. Amer. Math. Soc., *4* (1953), 203–212.

10. T. Motzkin and O. Taussky, *Pairs of matrices with property L. II*, Trans. Amer. Math. Soc., *80* (1955), 387–401.

11. A. Rosenberg and D. Zelinsky, *On Nakayama's extension of the $x^{n(x)} = x$ theorem*, Proc. Amer. Math. Soc., *5* (1954), 484–486.

Trinity College, Cambridge

RADICALS OF POLYNOMIAL RINGS

S. A. AMITSUR

Introduction. Let R be a ring and let $R[x]$ be the ring of all polynomials in a commutative indeterminate x over R. Let $J(R)$ denote the Jacobson radical **(5)** of the ring R and let $L(R)$ be the lower radical **(4)** of R. The main object of the present note is to determine the radicals $J(R[x])$ and $L(R[x])$. The Jacobson radical $J(R[x])$ is shown to be a polynomial ring $N[x]$ over a nil ideal N of R and the lower radical $L(R[x])$ is the polynomial ring $L(R)[x]$. A partial result of the first case and a parallel solution to the second case have been obtained also independently and by different methods by N. H. McCoy (simultaneously with the author).

The present method of attacking these problems can be applied to many other radicals arising from π-properties **(1)** of rings. Let $\pi(R)$ denote the π-radical of a ring R. $J(R)$ is an example of a radical satisfying $\pi(R[x]) = P[x]$ where $P = \pi(R[x]) \cap R$, and $L(R)$ represents a class of radicals satisfying $\pi(R[x]) = \pi(R)[x]$. The results obtained can be easily extended to polynomials in any number of variables.

It is shown that $J(R[x]) = N[x]$ where N is a nil ideal in R. Snapper, who studied the Jacobson radical of polynomial rings over commutative rings R, has shown **(7)** that N is the *maximal* nil ideal in R. The extension of this result to arbitrary rings seems to be very difficult. Though we verify this fact for algebras over non-denumerable fields, the general problem of determining the ideal N remains open.

1. The Jacobson radical

LEMMA 1J: *Let* $N = J(R[x]) \cap R$, *then* $J(R[x]) \neq 0$ *implies* $N \neq 0$.

The proof of this Lemma, which is a keystone in the extension of the results on the Jacobson radical to arbitrary radicals, seems to be rather elementary if R is an algebra over an infinite field, or if R is of characteristic zero; but the proof is far more complicated in the general case.

Recall that the Jacobson radical is a radical of the type dealt in **(1)**. In particular, it follows by Corollary 1.1 of **(1)** that $J(R[x])$ remains invariant under the automorphisms of $R[x]$. For example, consider the automorphism[1]: $f(x) \rightarrow f(x+1)$ of $R[x]$, or more generally the automorphism: $f(x) \rightarrow f(x+\lambda)$, where λ is an endomorphism of the additive group of R satisfying $\lambda(ab) = (\lambda a)b = a(\lambda b)$ for $a, b \in R$.

Put $J = J(R[x])$. If the Lemma is not true, then we have a case where $J \neq 0$ but $J \cap R = N = 0$. Let $f(x)$ be a non zero polynomial of minimum

Received July 25, 1955.

[1]One does not need to assume that R contains a unit.

degree belonging to J. By the previous remarks it follows also that $f(x+1) \in J$. Hence $f_0(x) = f(x+1) - f(x) \in J$ since the degree of $f_0(x)$ is less than that of $f(x)$. The minimality of the latter implies that $f_0(x) = 0$. Thus $f(x+1) = f(x)$.

If R is of characteristic zero, then one readily verifies that $f(x+1) = f(x)$ can hold only if $f(x) = a \in R$. Thus $0 \neq a \in J \cap R$ which is a contradiction. Another immediate contradiction is readily obtained if R is assumed to be an algebra over an infinite field F. Indeed, the preceding arguments can be applied as well to the automorphisms: $g(x) \to g(x+\lambda)$, $\lambda \in F$, of $R[x]$. This yields that $f(x+\lambda) = f(x)$ for all $\lambda \in F$. Since F is infinite, the last relation implies that $f(x) = a \in R$, and thus $0 \neq a \in N$ which contradicts the assumption $N = 0$.

In order to obtain a contradiction in the general case we have to make some additional remarks: Let p be a prime number and let R_p be the set of all elements of R which are of characteristic p. Note that R is an ideal in R and, therefore $R_p[x]$ is an ideal in $R[x]$. We may assume that $f(x) \in R_p[x]$. Indeed, let $f(x) = a_0x^n + \ldots + a_n$. Since $N = 0$ it follows that $n > 1$. Hence, since $f(x+1) - f(x) = na_0x^{n-1} + \ldots = 0$ we obtain that $na_0 = 0$. Let m be the minimal integer satisfying $ma_0 = 0$ and let p be a prime dividing m. Thus $(m/p)a_0 \neq 0$ and clearly $(m/p)a_0 \in R_p$. We may replace $f(x)$ by the polynomial $(m/p)f(x)$ which belongs also to J and which is of the same degree as $f(x)$. Namely, we suppose that the highest coefficient of $f(x)$ belongs to R_p. Now $pf(x) \in J$ and its degree is smaller than the degree of $f(x)$, hence the minimality of the latter yields $pf(x) = 0$, that is, $f(x) \in R_p[x]$.

Next we show that if a polynomial $g(x) \in R_p[x]$ satisfies $g(x+1) = g(x)$ then $g(x) = h(x^p - x)$ is a polynomial in $x^p - x$ with coefficients in R_p. The proof is carried out by induction on the degree of $g(x)$. First, let $g(x)$ be a polynomial of degree $k < p$. Since $g(x+1) = g(x)$ it follows that $g(x+\nu) = g(x)$ for all integers ν. Now,

$$g(x+\nu) = g(\nu) + xg_1(\nu) + \ldots + x^m g_m(\nu) = b_0 + xb_1 + \ldots + x^m b_m = g(x).$$

Hence, $g(\nu) = b_0$ for all integers ν. Clearly, R_p is an algebra over the finite field $GF[p]$ of p elements. Thus we obtained that $g(x) - b_0$ vanishes for p elements of the field $GF[p]$. Since the degree of $g(x) - b_0$ is less than p, it follows that $g(x) - b_0 = 0$, that is, $g(x) = b_0 \in R_p$.

Let $g(x)$ be a polynomial of arbitrary degree,[2] then $g(x) = h(x)(x^p - x) + k(x)$, with the degree of $k(x) < p$. Hence, $g(x+1) = h(x+1)(x^p - x) + k(x+1)$ and since $g(x+1) = g(x)$ we obtain

$$[h(x+1) - h(x)](x^p - x) = k(x) - k(x+1).$$

The degree of the right-hand side of the last equality is less than p and the degree of the left-hand side, if not zero, is $\geq p$. It follows, therefore, that

[2]This holds in the ring $R^*{}_p[x]$, where R^* is obtained by adjoining a unit to R, but, clearly, h(x), $g(x)$ and $k(x)$ belong to $R_p[x]$.

$k(x) = k(x + 1)$ and $h(x + 1) = h(x)$. By the previous case we know that $k(x) = k_0 \in R_p$ and by induction it follows that $h(x) = h_0(x^p - x)$. Thus $g(x) = h_0(x^p - x)(x^p - x) + k_0$ is a polynomial in $R_p[x^p - x]$.

The last preparatory remark we need before completing the proof of the Lemma is to the effect that if a polynomial $h(x^p - x)$ belongs to the Jacobson radical of $R_p[x]$, then it belongs also to the Jacobson radical of $R_p[x^p - x]$. Indeed, let $k(x) \in h(x^p - x) R_p[x^p - x]$, then $k(x + 1) = k(x)$. Clearly $k(x)$ belongs to the Jacobson radical of $R_p[x]$ hence its quasi-inverse $k'(x)$ is uniquely determined. The quasi-inverse of $k(x + 1)$ is readily seen to be $k'(x + 1)$; hence $k(x) = k(x + 1)$ implies that $k'(x + 1) = k'(x)$. Consequently $k'(x) \in R_p[x^p - x]$. This proves that the right ideal $h(x^p - x) R_p[x^p - x]$ is quasi-regular in $R_p[x^p - x]$. Thus $h(x^p - x)$ belongs to the Jacobson radical of $R_p[x^p - x]$.

We turn now to the proof of the Lemma. Since $f(x) \in J \cap R_p[x]$ and $R_p[x]$ is an ideal in $R[x]$, it follows that $f(x) \in J(R_p[x]) = R_p[x] \cap J$. It was shown that $f(x + 1) = f(x)$, hence $f(x) \in R_p[x^p - x]$. Thus by the previous remarks, it follows that $f(x) = g(x^p - x)$ and $g(x^p - x)$ belongs to the Jacobson radical $J(R_p[x^p - x])$. The mapping $h(x) \to h(x^p - x)$ determines an automorphism between $R_p[x]$ and $R_p[x^p - x]$. It follows now, by Theorem 1.7 of **(1)**, that $J(R_p[x])$ is the image of $J(R_p[x^p - x])$. In particular, it follows that since $g(x^p - x) = f(x) \in J(R_p[x^p - x])$, $g(x) \in J(R_p[x])$. But $g(x)$ is of lower degree[3] than $f(x)$; hence, $g(x) \in J(R_p[x]) = J \cap R_p[x]$ implies that $g(x) \in J$, which contradicts the minimality of $f(x)$. This completes the proof of the Lemma.

LEMMA 2J: *$J(R[x]) = N[x]$, where $N = J(R[x]) \cap R$.*

Indeed, since $N \subseteq J = J(R[x])$, it follows that $N[x]\, R[x] \subseteq NR[x] \subseteq J$. Hence $N[x] \subseteq J$. Consider the homomorphism: $R[x] \to R[x]/N[x]$. The kernel of this homomorphism is $N[x] \subseteq J$. It follows, therefore, by Theorem 1.7 of **(1)** that $J(R[x]/N[x]) = J/N[x]$. Let $\bar{R} = R/N$, then $R[x]/N[x] \cong \bar{R}[x]$. Now:

$$\begin{aligned} J(\bar{R}[x]) \cap \bar{R} \cong J/N[x] \cap (R, N[x])/N[x] &= (J \cap (R, N[x]))/N[x] \\ &= (J \cap R, N[x])/N[x] = (N, N[x])/N[x] = \bar{O} \end{aligned}$$

since $J \supseteq N[x]$. Hence, Lemma 1 implies that $\bar{O} = J(\bar{R}[x]) = J/N[x]$. Consequently, $J = N[x]$, as required.

It remains now to determine the structure of the ideal N.

LEMMA 3J: *N is a nil ideal in R.*

Clearly, N is an ideal in R. Let $r \in N \subseteq J$, then $r.rx = r^2x \in J$. Let $q(x)$ be the quasi-inverse of r^2x, that is, $q(x) + r^2x + q(x)r^2x = 0$. In other words

$$q(x) = -r^2x - q(x)r^2x.$$

[3] This is true since the degree of $f(x)$ is >1.

Substitute $q(x)$ on the right by the whole expression of the right-hand side of this equality. Repeating this process yields

$$q(x) = -r^2x + (r^2x)^2 + \ldots + (-1)^n(r^2x)^n + (-1)^{n+1}(r^2x)^{n+1} + (-1)^{n+1}q(x)(r^2x)^{n+1}.$$

Choose n to be greater than the degree of $q(x)$. Equating the coefficient of x^n on both sides yields that $r^{2n} = 0$. This proves that N is a nil ideal.

Levitzki's locally nilpotent radical $s\sigma(R)$ of a ring R is defined **(2**, p. 130**)** as the maximal ideal of R with the property that its finitely generated subrings are nilpotent. One readily observes that the polynomial ring $s\sigma(R)[x]$ is a nil ideal and, therefore, it also is quasi-regular. Consequently $s\sigma(R)[x] \subseteq J$, and thus $s\sigma(R) \subseteq N$. Summarizing the results obtained, we have

THEOREM 1. *$J(R[x]) = N[x]$ where $N = J(R[x]) \cap R$ is a nil ideal containing the locally nilpotent radical $s\sigma(R)$ of R.*

Remark. If R is commutative, or more generally satisfies a polynomial identity, then it is known **(6)** that the nil ideals of R are locally nilpoint ideals. Thus in this case $N \subseteq s\sigma(R)$, and therefore, $J(R[x])$ is a nil ideal and N is the maximal nil ideal of R.

We restrict ourselves now to the case where R is an algebra over an infinite field F. An ideal I in an algebra R is called an LBI-ideal **(3)** if I is a nil ideal and every finitely generated *submodule* of I is of bounded index. One readily observes that if $f(x) \in I[x]$, where I is an LBI-ideal, then $f(x)$ is nilpotent and its index is bounded by the index of the submodule of I generated by the coefficient of $f(x)$. Thus, $I[x] \subseteq J$. The maximal LBI-ideal, $LBI(R)$, of R is known **(3)** as the LBI-radical of R. Hence, the preceding arguments yield, in view of the fact that $LBI(R) \supseteq s\sigma(R)$, that:

COROLLARY. *$N \supseteq LBI(R) \supseteq s\sigma(R)$:*

It was shown in **(3)** that if R is an algebra over a *non denumerable* field F, then every nil ideal in R is an LBI-ideal. Consequently, for such algebras $LBI(R) \supseteq N$, which in particular implies that $N[x]$ is a nil ideal. Since the nil ideals are quasi-regular, it follows that:

THEOREM 2. *If R is an algebra over a non denumerable field F, then the Jacobson radical $J(R[x]) = N[x]$ is the maximal nil ideal of $R[x]$, and N is the maximal nil ideal of R.*

One conjectures that in all cases $J(R[x])$ is the maximal nil ideal of R. [x] This would follow immediately if one could supply a positive answer to the still-open problem of Levitzki which requires to show that every nil ring is locally nilpotent, since in that case $N = s\sigma(R)$ will hold for every ring.

2. The lower radical. Let $L(R)$ denote the lower radical of the ring R. From the results of **(2**, Corollary 2.2**)**, we know that the lower radical arises

from a property L of rings. Recall that a ring R is an L-ring if every non zero homomorphic image of R contains non zero nilpotent ideals. The property L satisfies the same requirements of **(1)** as the property of quasi-regularity: namely, $L(=\sigma^*$ in the notations of **(1)**) is an SRZ-property of rings. We have also in this case:

LEMMA 1L: *Let* $L = L(R[x]) \cap R$; *then* $L(R[x]) \neq 0$ *implies* $L \neq 0$.

LEMMA 2L: $L(R[x]) = L[x]$.

The proof of the two Lemmas follows in parallel lines the proof of Lemma 1J and Lemma 2J, except that at one place in the proof of Lemma 1J we have used the definition of quasi-regularity and not the general requirements of an SR property. The proof of this point for the lower radical is what remains to complete the proof of the present two Lemmas. That is: we have only to show that "if $f(x) \in R_p[x^p - x]$ belongs to the lower radical of $R_p[x]$, then it belongs also to the lower radical of $R_p[x^p - x]$." Indeed, the ideal generated by $f(x)$ in $R_p[x^p - x]$ is a subring of the ideal generated by $f(x)$ in $R_p[x]$. The latter is an L-ideal, since $f(x) \in L(R_p[x])$. By Corollary 2.2 of **(2)**, it follows that subrings of L-rings are L-rings; hence the ideal generated by $f(x)$ in $R_p[x^p - x]$ is also an L-ideal. Consequently, $f(x) \in L(R_p[x^p - x])$. This completes the proof of Lemma 1L and, therefore, also of Lemma 2L.

In parallel to Lemma 3J, one has to characterize the ideal L. In the present case we can show that $L = L(R)$.

THEOREM 3. $L(R[x]) = L(R)[x]$.

Indeed, since $L = L(R[x]) \cap R \subseteq L(R[x])$ and L is an ideal in R, it follows by Corollary 2.2 of **(2)** that L is an L-ideal in R. Hence $L \subseteq L(R)$. The converse $L(R) \subseteq L$ will follow immediately from the following:

LEMMA 3L. *If S is an L-ring then* $S[x]$ *is also an L-ring.*

Indeed, let $S[x] \rightarrow \overline{S[x]}$ be a homomorphism of $S[x]$ onto a ring $\overline{S[x]}$. This homomorphism induces a homomorphism of S onto a ring $\bar{S} \subseteq \overline{S[x]}$. If $\bar{x}$ denotes the image of x, then clearly $\overline{S[x]} = \bar{S}[\bar{x}]$. Thus if $\overline{S[x]} \neq 0$ then $\bar{S} \neq 0$. Since S is an L-ring, $\bar{S}$ contains a non zero nilpotent ideal $\bar{Q}$. Consequently, $\bar{Q}[\bar{x}]$ is a nilpotent ideal of $\overline{S[x]}$, which proves that $S[x]$ is also a an L- ring.

To complete the proof of Theorem 3, we note that Lemma 3L implies that $L(R)[x]$ is an L-ideal. Hence $L(R)[x] \subseteq L(R[x]) = L[x]$. Thus $L(R) \subseteq L$.

3. Infinite sets of indeterminates. Let $R[x_\alpha]$ be the ring of all polynomials in a set of α indeterminates $\{x_i\}$ where α is any cardinal number. A simple induction procedure, or a proof similar to that of Lemma 2J, yields

THEOREM 4 (a). $J(R[x_\alpha]) = N_\alpha[x_\alpha]$ *where* $N_\alpha = J(R[x_\alpha]) \cap R$ *is a nil ideal and* $N_\beta \supseteq N_\alpha \supseteq s\sigma(R)$ *for all* $\beta < \alpha$.

(b) $L(R[x_\alpha]) = L(R)[x_\alpha]$.

Furthermore, we have

THEOREM 5. *Let α be an infinite cardinal, then $J(R[x_\alpha]) = N_\alpha[x_\alpha]$ is the maximal nil ideal of $R[x_\alpha]$ and $N_\alpha = N_\beta$ for all $\beta > \alpha$. If R is an algebra over an infinite field, then $N_\alpha = LBI(R)$.*

Let x_1 be an indeterminate of the set $\{x_i\}$ and let $\{x_i'\}$ denote the rest of the indeterminates. Since α is not finite, the sets $\{x_i\}$ and $\{x_i'\}$ have the same cardinal number. Hence $J(R[x_i']) = N_\alpha[x_i]$ and $J(R[x_i]) = N_\alpha[x_i]$. Clearly $R[x_i] = R'[x_1]$ where $R' = R[x_i']$. It follows now by Theorem 1, that

$$J(R'[x_1]) = N'[x_1], \quad N' = R' \cap J(R'[x_1]).$$

Since $J(R'[x_1]) = J(R[x_i]) = N_\alpha[x_i]$, it follows that $N' = N_\alpha[x_i']$. By Theorem 1 it follows that N' is a nil ideal. Since $\{x_i\}$ and $\{x_i'\}$ are of the same cardinal number, one obtains $N_\alpha[x_i'] \cong N_\alpha[x_i]$. Consequently, $N_\alpha[x_i]$ is a nil ideal; thus $J(R[x_\alpha])$ is a nil ideal and, therefore, it is the maximal nil ideal of $R[x_\alpha]$.

Let $\{x_i\}$ be a set of indeterminates of cardinality $\alpha > N_0$ and let $\{y_j\}$ be a finite set of new indeterminates. Since the cardinality of the set $\{x_i\}$ and $\{x_i, y_j\}$ is α, we have $J(R[x_i, y_j]) = N_\alpha[x_i, y_j]$, $J(R[x_i]) = N_\alpha[x_i]$ where $N_\alpha = R \cap J(R[x_i]) = R \cap J(R[x_i, y_j])$. By the preceding proof it follows that $N_\alpha[x_i, y_j]$ is a nil ideal. Hence, the ring of all polynomials over $N_\alpha[x_i]$ in any number (finite or non finite) of indeterminates is a nil ring. Clearly, the non finite case can be reduced to the finite case which has just been proved.

Now let $\beta > \alpha$ and let $\{x_i\}$ be a set of indeterminates of cardinality α and $\{x_i, z_j\}$ a set of indeterminates of cardinality β. By the previous remark it follows that $N_\alpha[x_i, z_j]$ is a nil ideal, hence $N_\alpha[x_i, z_j] \subseteq J(R[x_i, z_j])$. On the other hand $J(R[x_i, z_j]) = N_\beta[x_i, z_j]$; hence, $N_\alpha \subseteq N_\beta$. Since $\beta > \alpha$, it follows by Theorem 4 that $N_\beta \subseteq N_\alpha$. Thus $N_\beta = N_\alpha$.

Let R be an algebra over an infinite field F and let $a_1, \ldots, a_n$ be a finite set of elements of N_0. Since α is an infinite ordinal, we have a finite set of indeterminates $x_1, \ldots, x_n \in \{x_i\}$ and thus, $a_1x_1 + \ldots + a_nx_n \in N_\alpha[x_\alpha]$. It follows by the previous result that $(a_1x_1 + \ldots + a_nx_n)^m = 0$ for some integer m. This immediately implies that the module generated by the set $(a_1, \ldots, a_n)$ contains nil elements of index $< m$. Consequently, $N_\alpha \subseteq LBI(R)$, and the fact that $N_\alpha \supseteq LBI(R)$ completes the proof of the theorem.

4. π-radicals. We follow in this section the notation of **(1)** and **(2)**.

The similarity between the proofs of Lemma 1J, 2J and Lemmas 1L, 2L exhibits the generality of the methods used. The only place where the explicit definitions of the quasi-regularity and the L-property were involved was in proving that if $f(x) \in R_p[x^p - x]$ belongs to the radical considered of $R_p[x]$, then it belongs also to the same type of radical of $R_p[x^p - x]$. The proof of this fact for the L-property uses only the fact that a subring of an L-ring is an L-ring. This condition for arbitrary properties π was denoted in **(1)** as (D_s). Thus we have:

LEMMA 4. *If π is an RZ-property satisfying (D_s) then $\pi(R_p[x]) \cap R_p[x^p - x] \subseteq \pi(R_p[x^p - x])$.*

The method used in proving Lemmas 1L, 2L and Lemmas 1J, 2J, yields also

THEOREM 1π. *If π is an RZ-property and R is an algebra over an infinite field or of characteristic zero, or π satisfies the condition that $\pi(R_p[x]) \cap R_p[x^p - x] \subseteq \pi(R_p[x^p - x])$ then: $\pi(R[x]) \neq 0$ implies that $\pi(R[x]) \cap R \neq 0$.*

THEOREM 2π. *If π and R are the same as in the preceding Lemma, then $\pi(R[x]) = P[x]$ where $P = \pi(R[x]) \cap R$.*

One readily verifies also, as in the proof of Theorem 3, that:

THEOREM 3π. *If π and R are as above and if π satisfies the condition that a polynomial ring $S[x]$ over a π-ring S is also a π-ring then $\pi(R[x]) = \pi(R)[x]$.*

Properties satisfying the conditions of Theorem 1π are readily seen to be nillity, locally finiteness and locally nilpotency. The latter satisfies also Theorem 3π.

REFERENCES

1. S. A. Amitsur, *A general theory of radicals* II, Amer. J. Math., *76* (1954), 100–125.
2. ———, *A general theory of radicals* III, Amer. J. Math., *76* (1954), 126–136.
3. ———, *Algebras over infinite fields*, Proc. Amer. Math. Soc. (to be published).
4. R. Baer, *Radical ideals*, Amer. J. Math., *69* (1947), 46–58.
5. N. Jacobson, *The radical and semi simplicity for arbitrary rings*, Amer. J. Math., *67* (1945), 300–320.
6. J. Levitzki, *A theorem on polynomial identities*, Proc. Amer. Math. Soc., *1* (1950), 331–333.
7. E. Snapper, *Completely primary rings* I, Ann. Math., *52* (1950), 666–693.

Hebrew University,
Jerusalem, Israel

CONICS AND ORTHOGONAL PROJECTIVITIES IN A FINITE PLANE

W. L. EDGE

1. Introduction. The ternary orthogonal group of projectivities over a finite field leaves a non-singular conic χ invariant, but the geometry consequent thereupon does not appear to have been investigated. The group is isomorphic to a binary group of fractional substitutions over the same field and this fact may, since these binary groups and their subgroups are so well known, have inhibited projects to embark on a detailed description of the geometry of the ternary group. While, however, one may concede that no new intrinsic knowledge of the group can be gained, different representations of the same abstract group are apt to portray some of its attributes from different aspects and to display in different settings interrelations among its properties; and if one recalls the situation in the real or complex field the incentive to initiate some investigation becomes compelling.

The representation, over the real or complex field, of the points of a line λ by those of a conic Γ is now commonplace and goes back at least as far as Hesse. The involutions of pairs of points, as well as harmonic sets, seem more appositely carried on Γ than on λ. The Pascal property of Γ is simply, in essence, a statement about three involutions having a pair in common; but although these involutions can be carried on any rational curve, and the Pascal property interpreted in that context, it will be generally agreed, and not merely on historical grounds, that the conic is the most appropriate setting for it. The representation, too, of harmonic pairs on λ as pairs on Γ whose joins are conjugate has its advantages, and no apology is needed for undertaking some account of the corresponding representation when the base field is neither the real nor the complex field but a Galois field.

The paper falls into three sections. In the first (§§2–9) the foundations of the figure are laid and its fundamental properties established. It is explained how the points of the plane fall into 3 disjoint classes according as they are exterior to, on, or interior to χ; this phenomenon is known **(10)**, but we proceed to discuss the pairing, on various lines, of conjugate points. This pairing is basically relevant, and the description of it has to take account of whether or not -1 is a square in the base field. The number of canonical triangles—triangles, that is, in reference to which χ is given by equating the unit quadratic form to zero—is calculated.

The second section (§§10–17) introduces the orthogonal group of projectivities and stresses the presence in it of involutions (of two kinds) and octahedral subgroups. The subgroup, of index 2, which subjects the points

Received October 6, 1955.

of χ to even permutations is the main focus of interest and a criterion is given for the octahedral subgroups to belong to it. They do, or do not, belong to it according as it permutes the canonical triangles intransitively or transitively. Other subgroups are found as the stabilisers of points in the plane.

The third section (§§18–32) is devoted to a detailed description of the geometry when the base field is $GF(p)$ and $p = 5, 7, 11$. For these values of p, but not for any higher values, the orthogonal group has a representation as a permutation group of degree p; such representations are found in the 3 planes. The geometry has many features of interest, such as the multiple perspectivities between certain pairs of canonical triangles and, when $p = 11$, the distribution of the points external to χ in sets of 6, the 15 joins of points of such a set being all skew to χ and concurrent in threes at 10 different points all internal to χ.

Conics and their Canonical Triangles

2. The number q of marks in a finite field F is always a power of a prime p. Every non-zero mark satisfies $x^{q-1} = 1$, and there always occur primitive marks of which no power lower than the $(q - 1)$th is 1. All the non-zero marks are powers of any primitive mark j. We suppose throughout that $p > 2$. Then j cannot be the square of any mark of F because, if $j = i^2$,

$$j^{\frac{1}{2}(q-1)} = i^{q-1} = 1,$$

contradicting the primitiveness of j. Nor can any odd power of j be a square; it is impossible to extract a square root of any odd power of j without enlarging F. All even powers of j, on the other hand, are clearly squares of marks of F. The non-zero marks are thus half of them squares and the other half non-squares.

The product and quotient of two non-squares are always squares.

Take, as an example, $q = p = 7$. We may label the marks

$$-3, -2, -1, 0, 1, 2, 3$$

and regard them as the residue classes to modulus 7. The primitive marks are 3 and -2. The squares are

$$1 = 3^6 = (-2)^6, \ -3 = 3^4 = (-2)^2, \ 2 = 3^2 = (-2)^4,$$

while the non-squares are

$$-1 = 3^3 = (-2)^3, \ 3 = 3^1 = (-2)^5, \ -2 = 3^5 = (-2)^1.$$

It is important, with a view to the geometry, to distinguish between fields wherein -1 is, or is not, a square. Since -1 is $j^{\frac{1}{2}(q-1)}$, this power of j not being 1 and yet a square root of 1,

$$-1 \text{ is a square whenever } q \equiv 1 \pmod 4,$$
$$-1 \text{ is a non-square whenever } q \equiv -1 \pmod 4.$$

3. The marks of F will serve as homogeneous coordinates of points and lines in a plane; each point or line answers to a vector of 3 components not all of which are 0. There are $q^3 - 1$ such vectors; but the $q - 1$ non-zero multiples of any given vector represent the same point, or line, so that the plane consists of

$$(q^3 - 1)/(q - 1) = q^2 + q + 1$$

points and of the same number of lines. When it is necessary to distinguish point and line the coordinates of a point may be written as a column vector and those of a line as a row vector. The number of points on a line and of lines through a point is

$$(q^2 - 1)/(q - 1) = q + 1.$$

4. We take for granted (**3**, p. 158) the fact that every non-singular conic can, by appropriate choice of the triangle of reference, be given by equating to zero the unit quadratic form $x^2 + y^2 + z^2$. Let us, before embarking on the main task of exposition, enquire into the geometrical significance of this canonical form. It certainly refers the conic to a self-polar triangle, but there is more to say than this because a conic has self-polar triangles which, when used as triangle of reference, do not permit this canonical form unless F is enlarged. The complete explanation has to take account of whether or not -1 is a square in F. The line $x = 0$ meets the conic where $y^2 + z^2 = 0$; if -1 is a square this yields two intersections, whereas if -1 is not a square there are no intersections; and the like occurs on $y = 0$ and on $z = 0$. If we describe any triangle which permits the canonical form $x^2 + y^2 + z^2 = 0$ as a *canonical triangle* and denote it by Δ then

if $q \equiv 1 \pmod 4$ the sides of any Δ are all *chords* of the conic,

if $q \equiv -1 \pmod 4$ the sides of any Δ are all *skew* to the conic.

The number, $q(q^2 - 1)/24$, of Δ is found below in §§7, 8.

Note that there is, on any side of any Δ, a unique pair of points that is both harmonic to the vertices of Δ and conjugate for the conic; on $x = 0$ this pair is given by $y^2 = z^2$, and that whether $x = 0$ is a chord or is skew to the conic. The three such pairs on the sides of a Δ are the vertices of a quadrilateral Q having Δ for its diagonal triangle; Δ and Q each determine the other uniquely. When Δ is the triangle of reference the sides of Q are

$$x + y + z = 0, \quad -x + y + z = 0, \quad x - y + z = 0, \quad x + y - z = 0.$$

5. Let χ denote the conic $x^2 + y^2 + z^2 = 0$. The polar of $P(\alpha, \beta, \gamma)$ is $\alpha x + \beta y + \gamma z = 0$, and passes through P if and only if P is on χ; χ is the aggregate of points that lie on their own polars.

If the polar of P passes through $P'(\alpha', \beta', \gamma')$ then $\alpha\alpha' + \beta\beta' + \gamma\gamma' = 0$ and the polar of P' passes through P; P, P' are then *conjugate* with respect to χ.

Does the polar of P meet χ? At least one coordinate of P, say γ, is not zero; then, for points of χ on the polar,

$$\gamma^2(x^2 + y^2) + (\alpha x + \beta y)^2 = 0,$$
$$(\gamma^2 + \alpha^2)x^2 + 2\alpha\beta xy + (\gamma^2 + \beta^2)y^2 = 0.$$

The discriminant of this quadratic is

$$\alpha^2\beta^2 - (\gamma^2 + \alpha^2)(\gamma^2 + \beta^2) \equiv -\gamma^2(\alpha^2 + \beta^2 + \gamma^2),$$

so that the quadratic has, or has not, roots in F according as $-\alpha^2 - \beta^2 - \gamma^2 \equiv -\Sigma\alpha^2$ is, or is not, a square.

If $-\Sigma\alpha^2$ is a square we call the polar a c-line or *chord*, and say that P is *external* to χ.

If $\Sigma\alpha^2 = 0$, P is *on* χ and the polar a t-line or *tangent*. It does not meet χ elsewhere.

If $-\Sigma\alpha^2$ is a non-square we call the polar an s-line; it is *skew* to χ, and P *internal* to χ.

This separation by a conic of the points of a plane into disjoint classes is noted by Qvist (**10**, pp. 9 and 19) but he does not proceed further, save to remark on the numbers of tangents through the points. If a tangent passes through P then the polar of P passes through the "contact" of the tangent, and conversely; hence there pass

two t-lines through any external point,
one t-line through any point of χ,
no t-line through any internal point.

We may call external points e-points, and internal points i-points .

Every t-line consists of $q + 1$ points; one is the contact, but the remaining q have all to be e. It follows, on polarising, that there are $q + 1$ lines through any point of χ, one line being the tangent and the remaining q all c. Hence, since q chords pass through any point of χ, χ consists of $q + 1$ points. Since the number of c-lines is $\frac{1}{2}q(q + 1)$ and of t-lines is $q + 1$, the number of s-lines is

$$q^2 + q + 1 - \tfrac{1}{2}q(q + 1) - (q + 1) = \tfrac{1}{2}q(q - 1),$$

and this must also be the number of i-points. Thus χ separates the $q^2 + q + 1$ points of the plane into disjoint batches of

$$\tfrac{1}{2}q(q + 1), \qquad q + 1, \qquad \tfrac{1}{2}q(q - 1)$$

and likewise the $q^2 + q + 1$ lines into these numbers of c-lines, t-lines, s-lines, respectively.

Any two t-lines intersect, and the $\frac{1}{2}q(q+1)$ e-points are thus accounted for.

6. Of the $q - 1$ points of a c-line, not on χ, half are i and half are e. For let $(\alpha_1, \beta_1, \gamma_1)$ and $(\alpha_2, \beta_2, \gamma_2)$ be any two distinct points of χ; any point on the

c-line which joins them is $(\alpha_1 + k\alpha_2, \beta_1 + k\beta_2, \gamma_1 + k\gamma_2)$ where k runs through the $q - 1$ non-zero marks of F. This point is e or i according as

$$-\Sigma(\alpha_1 + k\alpha_2)^2 \equiv -2k\Sigma\alpha_1\alpha_2$$

is, or is not, a square; it cannot be zero since the point is not on χ. But when k runs through the $q - 1$ non-zero marks, $-2k\Sigma\alpha_1\alpha_2$ does likewise, and since, of these marks, $\frac{1}{2}(q - 1)$ are squares and the others not, it follows that, of the $q - 1$ points of the c-line not on χ, $\frac{1}{2}(q - 1)$ are e and the others i. It follows too, on polarising, that through each e-point there pass, with 2 t-lines, $\frac{1}{2}(q - 1)$ c-lines and $\frac{1}{2}(q - 1)$ s-lines.

Since there are $\frac{1}{2}(q - 1)$ i-points on each of the $\frac{1}{2}q(q + 1)$ c-lines there pass, through each i-point,

$$\tfrac{1}{2}q(q + 1) \,.\, \tfrac{1}{2}(q - 1)/\tfrac{1}{2}q(q - 1) = \tfrac{1}{2}(q + 1)$$

c-lines, and so $\frac{1}{2}(q + 1)$ s-lines too. Polarisation then discloses that, of the $q + 1$ points on any s-line, half are e and half i.

7. Call two points, neither of them on χ, *similar* if they are either both e or both i; otherwise *opposite*.

Consider the pairing, as conjugate to one another, of the $q - 1$ points on c that are not on χ. Any conjugate pair is given by

$$(\alpha_1 \pm k\alpha_2, \quad \beta_1 \pm k\beta_2, \quad \gamma_1 \pm k\gamma_2)$$

for some non-zero k. Now the marks $\pm 2k\Sigma\alpha_1\alpha_2$ are both squares or both non-squares if -1 is a square, whereas if -1 is not a square one of the two marks is a square and the other not. Hence, for -1 not a square, the conjugates of the $\frac{1}{2}(q - 1)$ e-points on c are the $\frac{1}{2}(q - 1)$ i-points on c; conjugate points on c are opposite. If, however, -1 is a square conjugate points on c are similar; the $\frac{1}{2}(q - 1)$ e-points consist of $\frac{1}{4}(q - 1)$ conjugate pairs, as do the $\frac{1}{2}(q - 1)$ i-points. Let, -1 being a square, the pole of c be e_0, and let e_1 and e_2 be any one of the $\frac{1}{4}(q - 1)$ pairs of conjugate e-points on c; then each vertex of the triangle $e_0e_1e_2$ is an e-point and the triangle, being self-polar for χ, is a canonical triangle Δ. Since we may choose c, with its pole, in $\frac{1}{2}q(q + 1)$ ways and, thereafter, take any of the $\frac{1}{4}(q - 1)$ conjugate pairs of e-points on c the number of Δ is, since each of its 3 sides may be used to begin its construction,

$$\tfrac{1}{2}q(q + 1) \,.\, \tfrac{1}{4}(q - 1) \,.\, \tfrac{1}{3} = q(q^2 - 1)/24,$$

and each e-point is a vertex of $\frac{1}{4}(q - 1)$ of them. The lowest value of q for which -1 is a square is 5; there are then 5Δ whose 15 vertices account for the 15 e-points just once. When $q = 9$ there are 30Δ, each of the $45e$-points being a vertex of 2 of them.

8. The relation between conjugate points on s can be deduced from that on c. Suppose that e_1 and e_2, two external points on s, are conjugate; their polars

c_1 and c_2 pass through e_2 and e_1, respectively and meet at i_0, the pole of s. Hence conjugate points on c are opposite and -1 is not a square. But if e_1 and i_2 are conjugate points on s their polars c_1 and s_2 pass through i_2 and e_1 respectively and meet at i_0; hence conjugate points on c are similar and -1 a square. It follows that, when -1 is a square, conjugate points on s are opposite; the conjugates of the $\frac{1}{2}(q+1)$ i-points are the $\frac{1}{2}(q+1)$ e-points. But if -1 is not a square conjugate points on s are similar; there are $\frac{1}{4}(q+1)$ conjugate pairs of i-points and $\frac{1}{4}(q+1)$ of e-points.

When -1 is not a square each Δ has s-lines for its sides and i-points for its vertices. In order to construct a Δ we may choose any one of the $\frac{1}{2}q(q-1)$ s-lines as a side, and thereafter any of the $\frac{1}{4}(q+1)$ pairs of conjugate i-points on it as vertices. Since the construction may set out from any of the 3 sides the number of Δ is

$$\tfrac{1}{2}q(q-1)\,.\,\tfrac{1}{4}(q+1)\,.\,\tfrac{1}{3} = q(q^2-1)/24,$$

and each i-point is a vertex of $\frac{1}{4}(q+1)$ of them. The lowest value of q for which -1 is a non-square is 3; there is then a unique Δ and its vertices are the only i-points in the plane. When $q = 7$ there are 14Δ, each of the 21 i-points being a vertex of 2 of them; when $q = 11$ there are 55Δ, each of the 55 i-points being a vertex of 3 of them.

9. Let ABC be any Δ and take e, distinct from B and C whether the vertices be e-points or i-points, on BC (there is no such e if $q = 5$). The t-lines through e are harmonic to eBC and eA, and harmonic inversions in vertices and opposite sides of ABC yield a second pair of t-lines whose intersection is the harmonic conjugate of e in regard to B and C. These 4 t-lines form a quadrilateral U with ABC as diagonal triangle; U is the same as Q if $p = 3$, though not otherwise.

When the vertices of Δ are e the e-points on BC afford $\frac{1}{4}(q-5)$ pairs harmonic to B and C; Δ gives rise to $\frac{1}{4}(q-5)$ U, of which it is the diagonal triangle, whose sides account for all $q+1$ t-lines save those 6 which pass 2 through each of A, B, C. Every Δ provides such a partitioning of the t-lines. The lowest relevant values of q are 9 (when the U are also Q) and 13.

When the vertices of Δ are i the partitioning of t-lines is simpler; each Δ gives rise to $\frac{1}{4}(q+1)$ U, the e-points on any s-line falling into $\frac{1}{4}(q+1)$ pairs harmonic to the vertices of any Δ of which this s-line is a side; these e-points can be paired not only in the involution I_0 of pairs conjugate for χ, but in $\frac{1}{4}(q+1)$ involutions I_k each having the vertices on s of a Δ for foci. No two of these $\frac{1}{4}(q+5)$ involutions are the same, and I_0 commutes with all the others since the foci of any of these form a pair of I_0. When $q = 7$ the pairing of the 4 e-points is as follows:

e_1, e'_1 and e_2, e'_2 in I_0;
e_1, e_2 and e'_1, e'_2 in I_1, with foci i_1, i'_1;
e_1, e'_2 and e'_1, e_2 in I_2, with foci i_2, i'_2.

Here i_2, i'_2 must be a pair of I_1; i_1, i'_1 a pair of I_2; not only do I_1 and I_2 commute with I_0, they commute with each other. Their product, in either order, is I_0 as is seen by observing the permutations imposed on the e when I_1 and I_2 act in succession.

The Ternary Orthogonal Group of Projectivities

10. Suppose now that a projectivity leaves χ invariant. It must permute the Δ among themselves, so that the sides $x = 0$, $y = 0$, $z = 0$ of any given Δ become the sides $\xi = 0$, $\eta = 0$, $\zeta = 0$ of (the same or) some other Δ. Here ξ, η, ζ are linearly independent linear forms in x, y, z; and since χ admits both the equations

$$x^2 + y^2 + z^2 = 0, \qquad \xi^2 + \eta^2 + \zeta^2 = 0,$$

the left-hand side of either equation is a scalar multiple of the left-hand side of the other. Thus

$$\mathbf{\Xi} = \begin{bmatrix} \xi \\ \eta \\ \zeta \end{bmatrix} = \mathbf{M}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{M}\,\mathbf{x}$$

where $\mathbf{M}$ is a three-rowed non-singular matrix whose elements are all in F, and

$$\mathbf{x}'\mathbf{x} = x^2 + y^2 + z^2 = \lambda(\xi^2 + \eta^2 + \zeta^2) = \lambda\mathbf{\Xi}'\mathbf{\Xi} = \lambda\mathbf{x}'\mathbf{M}'\mathbf{M}\mathbf{x},$$

so that

10.1 $$\lambda\mathbf{M}'\mathbf{M} = \mathbf{I},$$

the unit matrix. Here λ is a mark of F; indeed it is a square because, on taking determinants in 10.1,

$$\lambda^3|\mathbf{M}|^2 = 1.$$

The projectivity is, however, unaffected if $\mathbf{M}$ is replaced by any scalar multiple of itself; if $\mathbf{H} = m^{-1}\mathbf{M}$ with m either square root of λ then, from 10.1,

$$\mathbf{H}'\mathbf{H} = \mathbf{I}.$$

Then $|\mathbf{H}|^2 = 1$, and we choose m to be that square root of λ for which $|\mathbf{H}|$ is $+1$; the projectivity is imposed by an orthogonal matrix of determinant $+1$. Conversely: this matrix is uniquely determined. For the only matrices which impose the same projectivity as $\mathbf{H}$ imposes are those of the form $\omega\mathbf{H}$ with ω a non-zero mark of F; the orthogonality condition demands that $\omega^2 = 1$ and the determinantal condition that $\omega^3 = 1$, which together require $\omega = 1$.

These projectivities, as likewise the unimodular orthogonal matrices that impose them, form a group $\Omega(3, q)$: the orthogonal group in 3 variables over F.

11. Note, in passing, the ***involutions*** in $\Omega(3, q)$, namely the harmonic inversions whose centre and axis are pole and polar for χ. Since the matrix imposing such an involution satisfies $\mathbf{H}^2 = \mathbf{I}$ as well as $\mathbf{H}'\mathbf{H} = \mathbf{I}$ it is symmetric as well as orthogonal. There are $\frac{1}{2}q(q+1)$ ***hyperbolic involutions*** whose centres are e-points and axes c-lines; the $q-1$ points of χ not on the axis are transposed in pairs. There are $\frac{1}{2}q(q-1)$ ***elliptic involutions*** whose centres are i-points and axes s-lines; the $q+1$ points of χ are transposed in pairs. One of the two consecutive integers $\frac{1}{2}(q \pm 1)$ is odd so that there are always in $\Omega(3, q)$ involutions that impose odd permutations on the points of χ—the hyperbolic ones if $q \equiv -1 \pmod 4$, the elliptic ones if $q \equiv 1 \pmod 4$.

12. The conditions, expressed by $\mathbf{H}'\mathbf{H} = \mathbf{I}$, for

$$\mathbf{H} = \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \beta_1 & \beta_2 & \beta_3 \\ \gamma_1 & \gamma_2 & \gamma_3 \end{bmatrix}$$

with all its elements in F, to be orthogonal are

12.1 $$\Sigma\alpha_1^2 = \Sigma\alpha_2^2 = \Sigma\alpha_3^2 = 1,$$

12.2 $$\Sigma\alpha_2\alpha_3 = \Sigma\alpha_3\alpha_1 = \Sigma\alpha_1\alpha_2 = 0.$$

These conditions can be interpreted geometrically when each column of $\mathbf{H}$ is regarded as the coordinate vector of a point of the plane; 12.2 then demands that the 3 points form a self-polar triangle for χ and 12.1 that their coordinate vectors be normalised. It is not possible to normalize any vector unless $\Sigma\alpha^2$ is a square; hence if -1 is a square internal points, and if -1 is a non-square external points, cannot have their coordinates normalised. But when it is possible to normalise a vector it admits two normalised forms $\pm(\alpha, \beta, \gamma)$.

If -1 is a square, each column of $\mathbf{H}$ is one of two normalised coordinate vectors of one of three mutually conjugate e-points; that is, the columns of $\mathbf{H}$ answer one to each vertex of a Δ. The vertices of Δ can be taken in any order and, with this order chosen, four of the eight combinations of sign are permitted by the stipulation that $|\mathbf{H}| = +1$. Hence the number of such orthogonal matrices is

$$\frac{q(q^2-1)}{24} \,.\, 3! \,.\, 4 = q(q^2-1).$$

If -1 is not a square, the calculation leads to the same result; it is governed by the columns representing vertices of a Δ and is not affected by these vertices being e or i. *The order of the group* $\Omega(3, q)$ *is* $q(q^2-1)$.

The triangle of reference is itself a Δ, and the 24 matrices, obtained from $\mathbf{I}$ by imposing the 3! permutations on its columns and using the 4 choices of sign permitted for each permutation, form that subgroup of $\Omega(3, q)$ for which the triangle of reference is invariant. It is an octahedral subgroup; indeed it acts as the symmetric group $\mathfrak{S}_4$ on the sides of the Q associated with the

triangle of reference, imposing all 4! permutations on them. When $q = 3$ this $\mathfrak{S}_4$ is the whole orthogonal group; otherwise it is one among $q(q^2-1)/24$ octahedral subgroups of $\Omega(3, q)$. The involutions in $\mathfrak{S}_4$ are those three whose centres are vertices of Δ and those six whose centres are vertices of Q. The former answer to the diagonal matrices diag$(1, -1, -1)$, diag$(-1, 1, -1)$, diag$(-1, -1, 1)$ and the latter to the matrices

$$\begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & -1 \\ 0 & -1 & 0 \\ -1 & 0 & 0 \end{bmatrix},$$
$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

13. $\Omega(3, q)$ acts as a permutation group on the $q + 1$ points of χ: should any of its operations subject these points to an odd permutation precisely one half of them must do so, and those operations that impose even permutations will then form a normal subgroup of index 2. We have, however, already noted the presence, in every group $\Omega(3, q)$, of involutions that impose odd permutations; hence $\Omega(3, q)$ has a normal subgroup $\Omega^+(3, q)$ of order $\frac{1}{2}q(q^2-1)$. Those involutions that belong to $\Omega^+(3, q)$ are the $\frac{1}{2}q(q+1)$ hyperbolic involutions if $q \equiv 1 \pmod 4$, whereas they are the $\frac{1}{2}q(q-1)$ elliptic evolutions if $q \equiv -1 \pmod 4$; in other words they are always those involutions whose centres are vertices of Δ.

14. There is a criterion which decides whether the octahedral subgroups of $\Omega(3, q)$ are also contained in $\Omega^+(3, q)$: they will not be so contained unless the vertices of Q are similar to those of Δ, for $\Omega^+(3, q)$ only contains either hyperbolic or elliptic involutions, never both. On the other hand, it does contain all the involutions of one of the two types. The test of $-\Sigma\alpha^2$ being, or not being, a square establishes that the vertices of Δ are similar to those of Q if, and only if, 2 is a square; for $-\Sigma\alpha^2$ is -1 at the vertices of the triangle of reference and -2 at those of its associated Q. When q is a prime p the similarity of the vertices requires that, in the common phraseology, 2 is a quadratic residue; this occurs **(11**, p. 110**)** whenever $p \equiv \pm 1 \pmod 8$, but not when $p \equiv \pm 3 \pmod 8$.

It is clear from §12 that $\Omega(3, q)$ permutes the Δ transitively; a given Δ is then invariant for

$$q(q^2-1) \div \frac{q(q^2-1)}{24} = 24$$

projectivities of $\Omega(3, q)$ and they form one of the octahedral subgroups. But if $\Omega^+(3, q)$ permutes the Δ transitively only 12 of its operations leave a given Δ invariant; they form a subgroup of index 2 in an octahedral group—a tetrahedral group that imposes the 12 even permutations on the sides of the associated Q. Should, therefore, $\Omega^+(3, q)$ contain the octahedral subgroups

it cannot act transitively on the Δ, which must then fall, under $\Omega^+(3, q)$, into two transitive sets of $q(q^2 - 1)/48$ each, sets which form two systems of imprimitivity for $\Omega(3, q)$ and which are transposed by any projectivity of $\Omega(3, q)$ that is outside $\Omega^+(3, q)$.

15. The rules (see §12) by which matrices of $\Omega(3, q)$ are formed show that the group is transitive not only on the Δ but also on those points that can serve as vertices of Δ; hence any such point B, and its polar b, are latent for a subgroup Ω_B, the stabiliser of B in $\Omega(3, q)$, of order

$$q(q^2 - 1) \div \tfrac{1}{2}q(q \pm 1) = 2(q \mp 1),$$

the upper or lower sign occurring according as B is e or i. Ω_B includes all the involutions whose centres are on b; they account for $q \mp 1$ of its operations and their matrices all have the coordinate vector of B latent, with multiplier -1. The other $q \mp 1$ operations of Ω_B form the group Ω_{B+} for which the coordinate vector of B is invariant, being associated with a latent root $+1$. Ω_{B+} is isomorphic to the binary orthogonal group which it induces on b, and we now show that it is cyclic. That it is abelian follows at once from the form of the 2-rowed orthogonal matrices of determinant 1 (**3**, p. 169); it can be asserted to be cyclic once the presence in it is detected of an operation whose period is the order of Ω_{B+}.

The binary orthogonal group consists of all matrices

$$\mathbf{U} = \begin{bmatrix} u & v \\ -v & u \end{bmatrix}$$

with $u^2 + v^2 = 1$ and both u, v belonging to $GF(q)$. Since $\mathbf{U}^2 = 2u\mathbf{U} - \mathbf{I}$ it follows (cf. **11**, p. 368) that, if h is a square root of -1,

$$2hv\mathbf{U}^n = \{(u + hv)^n - (u - hv)^n\}\mathbf{U} - \{(u + hv)^{n-1} - (u - hv)^{n-1}\}\mathbf{I},$$

and then that $\mathbf{U}^n = \mathbf{I}$ if, and only if,

$$(u + hv)^n = (u - hv)^n = 1.$$

Should B be an e then $h \in GF(q)$ and every $\mathbf{U}$ satisfies $\mathbf{U}^{q-1} = \mathbf{I}$; in order to find $\mathbf{U}$ with period $q - 1$ it is only necessary to choose $u + hv$, and therewith its reciprocal $u - hv$, to be a primitive mark.

If, however, B is an i then $u \pm hv$ do not belong to $GF(q)$ but to a quadratic extension $GF(q^2)$; they are conjugate marks therein, each the qth power of the other. Hence

$$(u + hv)^{q+1} = (u + hv)(u + hv)^q = (u + hv)(u - hv) = 1,$$

and every $\mathbf{U}$ satisfies $\mathbf{U}^{q+1} = \mathbf{I}$. A matrix of period $q + 1$ is found by choosing $u + hv$ in $GF(q^2)$ to be a primitive root of $x^{q+1} = 1$. The mark h, having served its purpose, falls out of the working and leaves only marks of $GF(q)$ in the final result.

Two examples may perhaps be given, with details of the calculations left out.

The quadratic $X^2 = X + 1$ is irreducible over $GF(7)$ and the adjunction of either root ζ extends the field to $GF(7^2)$. A primitive root of $x^8 = 1$ in $GF(7^2)$ is ζ^2, and so we take

$$u + \zeta^4 v = \zeta^2, \qquad u - \zeta^4 v = \zeta^{-2},$$

which give $u = v = -2$;

$$\mathbf{U} = \begin{bmatrix} -2 & -2 \\ 2 & -2 \end{bmatrix} \text{ has period 8 over } GF(7).$$

The quadratic $X^2 = X - 1$ is irreducible over $GF(11)$ and the adjunction of either root θ extends the field to $GF(11^2)$. A primitive root of $x^{12} = 1$ in $GF(11^2)$ is $2\theta + 2$ and we take

$$u + (4\theta - 2)v = 2\theta + 2, \qquad u - (4\theta - 2)v = -2\theta + 4,$$

which give $u = 3$, $v = -5$;

$$\mathbf{U} = \begin{bmatrix} 3 & -5 \\ 5 & 3 \end{bmatrix} \text{ has period 12 over } GF(11).$$

The operation of period 2 in Ω_{B+} is manifestly the involution with B as centre and this, since B is vertex of a Δ, belongs to $\Omega^+(3, q)$. But not all operations of Ω_{B+} so belong, and hence only half of them will do so. For Ω_B certainly contains operations outside $\Omega^+(3, q)$, namely those centred at $\frac{1}{2}(q \mp 1)$ points on b opposite to vertices of Δ; and when *all* involutions centred on b are jettisoned from Ω_B to leave Ω_{B+} only half these are outside $\Omega^+(3, q)$ and so not all the $q \mp 1$ operations of Ω_B outside $\Omega^+(3, q)$ are rejected. The consequence is that $\Omega^+(3, q)$ has subgroups Ω_B^+ of order $q \mp 1$ and cyclic subgroups Ω_{B+}^+ of order $\frac{1}{2}(q \mp 1)$, and is transitive on vertices of Δ.

It is perhaps not superfluous to remark that, as the involution centred at B imposes the identity projectivity on the points of b, the groups of projectivities on b are of orders one half those of the groups of orthogonal matrices; Ω_{B+}^+ imposes $\frac{1}{4}(q \mp 1)$ projectivities on b and is in (2, 1) homomorphism with this latter group.

16. There is a corresponding discussion for points D, opposite to vertices of Δ, and their polars d; $\Omega(3, q)$ has subgroups Ω_D of order $2(q \pm 1)$ and cyclic subgroups Ω_{D+} of order $q \pm 1$; $\Omega^+(3, q)$ has subgroups Ω_D^+ of order $q \pm 1$ and cyclic subgroups Ω_{D+}^+ of order $\frac{1}{2}(q \pm 1)$. There is a difference in that the "restriction" of the ternary quadratic form to the line d does not have the canonical form $\xi_1^2 + \xi_2^2$, that it had on the side of a Δ, but $\xi_1^2 + \nu\xi_2^2$ where ν is any fixed non-square of $GF(q)$. When the coördinates are transformed to correspond thereto the binary orthogonal projectivities answer to matrices (**3**, p. 161)

$$\mathbf{U} = \begin{bmatrix} u & v \\ -\nu v & u \end{bmatrix},$$

wherein $u^2 + \nu v^2 = 1$; but these also satisfy $\mathbf{U}^2 = 2u\mathbf{U} - \mathbf{I}$.

17. The group $\Omega^+(3, q)$ is (**3**, p. 164) isomorphic to the linear fractional group $LF(2, q)$, and the subgroups that have just been obtained in the orthogonal representation were found in the linear fractional representation by Serret for the case $q = p$ (**11**, pp. 375, 379, 380; results which were given also in the earlier editions of this treatise) and by Dickson for q a power of a prime (**3**, pp. 263–4). The 3 involutions centred at the vertices of any Δ are mutually commutative and form, with the identity, a 4-group; the orthogonal representation thus discloses the $q(q^2 - 1)/24$ 4-groups in $\Omega^+(3, q)$ at a glance. They are, of course, well known in the linear fractional representation (**3**, p. 268; **1**, p. 444). Serret also obtained the p^2 involutions of the group of linear fractional transformations, pointing out (**11**, p. 382) that the number in $LF(2, p)$ is $\frac{1}{2}p(p + 1)$ or $\frac{1}{2}p(p - 1)$ according as $p \equiv 1$ or -1 (mod 4).

The Details of the Geometry over the Smaller Fields

18. The remaining sections of the paper are given to describing the geometry for the smaller fields $q = 3, 5, 7, 11$. The figure for $q = 3$ has been described elsewhere (**4**); the Δ and Q therein are unique, and the sides of Q are the tangents at the 4 points of χ. $\Omega(3, 3)$ is the octahedral, $\Omega^+(3, 3)$ the tetrahedral, group and the points of χ and sides of Q undergo the corresponding permutations. The subgroups Ω_B, one for each vertex of Δ, are the dihedral subgroups of order 8; Ω_{B+} the cyclic subgroups of order 4. There is only a single Ω_B^+, namely the 4-group that is a common subgroup of the 3 dihedral Ω_B, but there are 3 cyclic groups Ω_{B+}^+ of order 2.

19. Some description has also been printed (**5**) of the figure for $q = 5$, although in quite a different context and using a different nomenclature. An account of this figure from the standpoint of the present enquiry is therefore given now. Each of the 15 c-lines is a side of one, and only one, Δ; and since no vertex e of any Δ lies on χ the c-lines through a point of χ belong one to each of the 5Δ. The 6 points of χ are separated by the sides of any Δ into 3 pairs of a *syntheme* (i.e., 3 pairs which together account for all 6 points) and the 5 synthemes, one arising from each Δ, constitute a *synthematic total* T (i.e, 5 synthemes which together account, by 3 pairs in each, for all 15 pairs of the 6 points).

Since there are 3 c-lines through each i-point the points of χ fall, in 10 distinct ways, into 3 pairs which, since their joins are concurrent, are in involution on χ. Each involution yields a syntheme, and the 10 synthemes so arising are those extraneous to T. We may say, with Clebsch, that the points of χ form a hexagon endowed 10 times over with the Brianchon property.

Clebsch (**2**, p. 336) establishes the existence of such hexagons in the real projective plane; their vertices are not then on a conic, neither will they be when we encounter such hexagons again below with $q = 11$. They arose when Clebsch mapped his 'diagonal' cubic surface on the plane, the surface itself

having arisen by making certain transformations of a quintic equation. The presence of such hexagons in the real plane is however visually obvious: they are provided by sections of the 6 diagonals of a regular icosahedron in Euclidean 3-space, the Brianchon points being the sections of the 10 joins of centroids of pairs of opposite faces. Some approach, whether deliberate or not, to this aspect of the matter is made by Klein (**8**, p. 218) but he does not appear to record that mere section is enough to provide the figure. The simplest section, by a plane perpendicular to a diagonal of the icosahedron, gives a hexagon consisting of the 5 vertices and the centre of a regular pentagon.

20. Take now any two Δ; call them Δ_1 and Δ_2. Label the points of χ A, B', C, A', B, C' so that

$$BC',\ CA',\ AB' \text{ are sides of } \Delta_1,$$
$$B'C,\ C'A,\ A'B \text{ are sides of } \Delta_2.$$

Since no e-point can lie on sides of more than one Δ, BC' and $B'C$ meet at a point i_1; the remaining c through i_1 is AA'. Hence

$$AA',\ BC',\ B'C \text{ are concurrent at } i_1,$$
$$BB',\ CA',\ C'A \text{ are concurrent at } i_2,$$
$$CC',\ AB',\ A'B \text{ are concurrent at } i_3;$$

moreover i_1, i_2, i_3 lie on the Pascal line s_0 of the hexagon $AB'CA'BC'$. Thus Δ_1 and Δ_2 are in *fourfold perspective*. Since they are both self-polar for χ, any axis of perspective is the polar of the corresponding centre of perspective; the four axes of perspective are AA', BB', CC', s_0. The first three of these are concurrent at i_0, the pole of s_0; they are the c-lines which pass through i_0 and are sides one of each of the Δ other than Δ_1 and Δ_2. Every pair of the 5Δ is in this relation of fourfold perspectivity, and each of the 10 s-lines plays the role of s_0 for one pair of Δ The s-lines are sides of Q, and each pair of Q share one s-line, namely the Pascal line of the hexagon of which their diagonal triangles provide alternate sides.

21. $\Omega(3, 5)$ is of order 120 and subjects the 5Δ to all 5! permutations; the octahedral subgroup for which one Δ is invariant subjects the other 4Δ to all 4! permutations. $\Omega^+(3, 5)$ is of order 60 and subjects the 5Δ to all even permutations; it has no octahedral subgroups and permutes the Δ transitively. The coordinate vectors of the 3 vertices of any Δ can therefore be displayed as columns of a matrix of $\Omega^+(3, 5)$; for instance thus:

$$21.1 \qquad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 2 & -1 & -1 \\ -1 & 1 & 2 \\ -1 & 2 & 1 \end{bmatrix}, \begin{bmatrix} 1 & -1 & 2 \\ -1 & 2 & -1 \\ 2 & -1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 2 & -1 \\ 2 & 1 & -1 \\ -1 & -1 & 2 \end{bmatrix}, \begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 1 \end{bmatrix}.$$

Indeed these matrices are symmetric and so, apart from **I**, represent involutions; the latent column vectors associated with the latent root $+1$ are found to be e, and so the involutions are hyperbolic and belong to $\Omega^+(3, 5)$.

Each matrix affords, by the 3! permutations of its columns and the 4 permissible signings for each permutation, 24 matrices of $\Omega(3, 5)$. If the permutations are restricted to be even, the 60 matrices so arising constitute $\Omega^+(3, 5)$; for the 12 arising from **I** form the tetrahedral subgroup of $\Omega^+(3, 5)$ for which the triangle of reference is invariant and these, when they postmultiply the other matrices of 21.1 impose even permutations on their columns.

The subgroups Ω_B are easily disposed of. If B is $y = z = 0$ the 8 matrices of Ω_B are

$$\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}, \begin{bmatrix}1&0&0\\0&-1&0\\0&0&-1\end{bmatrix}, \begin{bmatrix}1&0&0\\0&0&1\\0&-1&0\end{bmatrix}, \begin{bmatrix}1&0&0\\0&0&-1\\0&1&0\end{bmatrix}, \begin{bmatrix}-1&0&0\\0&1&0\\0&0&-1\end{bmatrix},$$

$$\begin{bmatrix}-1&0&0\\0&-1&0\\0&0&1\end{bmatrix}, \begin{bmatrix}-1&0&0\\0&0&1\\0&1&0\end{bmatrix}, \begin{bmatrix}-1&0&0\\0&0&-1\\0&-1&0\end{bmatrix},$$

the first 4 of which constitute Ω_{B+}, the first 2 Ω^+_{B+}. The 4 diagonal matrices constitute Ω^+_B, and the same subgroup Ω^+_B arises for the 3 vertices of any Δ. These are the 5 4-groups in $\Omega^+(3, 5)$. The 15 cyclic subgroups Ω^+_{B+} of order 2 are of course generated by the hyperbolic involutions.

22. The presence of axes of perspectivity is an immediate consequence of the factorisations of linear combinations of products of sides of any two Δ. For instance: the first and fifth of the matrices 21.1 provide, over $GF(5)$, the identities

$$(2x + y + z)(x + 2y + z)(x + y + 2z) + xyz \equiv (2x - y - z)(2y - z - x)(2z - x - y)$$

$$(2x + y + z)(x + 2y + z)(x + y + 2z) - xyz \equiv 2(x + y + z)(x^2 + y^2 + z^2),$$

and 9 other pairs of identities arise from these by applying the orthogonal transformations.

23. Suppose now that $q = 7$. All Δ and Q have i-points for their vertices, but whereas the sides of Δ are s those of Q are c. $\Omega^+(3, 7)$ has 14 octahedral subgroups, each acting as the symmetric group of 4! permutations of the sides of a Q.

Consider, for the moment, some one Δ. There are operations of $\Omega^+(3, 7)$ that permute its vertices cyclically; they leave one side c of the associated Q invariant while cyclically permuting the 3 i-points thereon. They cannot, being of odd period, transpose the two points of χ on c; nor can they, as not

imposing the identity projectivity on c, have any other latent point on c than these two; hence the 3 e-points on c are also permuted cyclically. Thus, on any c-line, the 3 i-points form an equianharmonic tetrad with either point of χ, as do likewise the 3 e-points. The two triads are analogous to a binary cubic and its cubic covariant, with their common Hessian pair.

24. The stabiliser Ω_B^+ of a given i-point, called for the moment B, in $\Omega^+(3, 7)$ is of order 8. There are 2 canonical triangles Δ, Δ' with B for vertex and these *are never transposed* by any operation of Ω_B^+. Δ is invariant for an octahedral subgroup of 24 operations of $\Omega^+(3, 7)$, and of these 8 leave B unaltered and so exhaust the stabiliser. The two octahedral subgroups associated with Δ and Δ' have this (dihedral) stabiliser in common. It follows that any operation of $\Omega(3, 7)$ that transposes Δ and Δ' lies outside $\Omega^+(3, 7)$. Any two Δ which share a vertex belong to different imprimitive systems.

It is then easy, starting from any one Δ (say the triangle of reference Δ_0), to obtain all the Δ and partition them into two sets of 7. There are 3, say Δ_1, Δ_2, Δ_3, which share a vertex with Δ_0 and so belong to the opposite set; each of them shares a vertex with two Δ other than Δ_0, and the 6Δ so arising all belong to the same set as Δ_0 and, indeed, complete it. We now display all 14Δ; each of the two horizontal strata consists of a set of 7 that are permuted transitively by $\Omega^+(3, 7)$. Each stratum is an imprimitive system for $\Omega(3, 7)$; whereas both strata are invariant for $\Omega^+(3, 7)$, they are transposed by those operations of $\Omega(3, 7)$ that lie outside $\Omega^+(3, 7)$.

$$
\begin{array}{|rrr|rrr|rrr|rrr|}
1 & 0 & 0 & 0 & -2 & -2 & 0 & 2 & -2 & 3 & 2 & -3 \\
0 & 1 & 0 & -2 & 3 & -3 & 2 & 3 & 3 & 2 & 0 & 2 \\
0 & 0 & 1 & -2 & -3 & 3 & -2 & 3 & 3 & -3 & 2 & 3 \\
\hline
-1 & 0 & 0 & 2 & 0 & 2 & -2 & 2 & 0 & 2 & -3 & -3 \\
0 & 2 & 2 & 0 & -1 & 0 & 2 & 2 & 0 & -3 & 2 & 3 \\
0 & 2 & -2 & 2 & 0 & -2 & 0 & 0 & -1 & -3 & 3 & 2
\end{array}
$$

$$
\begin{array}{rrr|rrr|rrr|}
3 & -2 & 3 & 3 & -3 & -2 & 3 & 3 & 2 \\
-2 & 0 & 2 & -3 & 3 & -2 & 3 & 3 & -2 \\
3 & 2 & 3 & -2 & -2 & 0 & 2 & -2 & 0 \\
\hline
2 & -3 & 3 & 2 & 3 & -3 & 2 & 3 & 3 \\
-3 & 2 & -3 & 3 & 2 & -3 & 3 & 2 & 3 \\
3 & -3 & 2 & -3 & -3 & 2 & 3 & 3 & 2
\end{array}
$$

Each square block provides, by permutations and signings of its columns, 24 matrices of $\Omega(3, 7)$; all 336 operations of the group are thus accounted for. The upper stratum provides, from its 7 blocks, the 168 matrices of $\Omega^+(3, 7)$; the unit matrix provides those 24 matrices for which the triangle of reference is invariant. The other octahedral subgroups occur when these 24 matrices are transformed, in the sense $\mathbf{HMH}^{-1}$, by those of the other 13 blocks.

25. Each block is symmetric and therefore the matrix of an involution save when it is the unit matrix. The involutions of $\Omega^+(3, 7)$ are the 21 elliptic ones; of these 9 are provided by the matrices given in §12. There are 12 others, of which 6 are furnished by the blocks precisely as displayed; the outstanding 6 are got from these by changing the signs of those 4 marks that occur in the same row or column as the zero in the diagonal, these changes neither altering the value of the determinant nor destroying the orthogonality. The 28 hyperbolic involutions must be imposed by symmetric matrices, orthogonal and of determinant $+1$, whose columns, either themselves or their negatives, occur in the lower stratum. All the assemblages

$$\begin{vmatrix} 2 & 2 \\ 2 & -2 \end{vmatrix} \begin{matrix} 2 & -2 \\ -2 & -2 \end{matrix} \begin{vmatrix} -2 & 2 \\ 2 & 2 \end{vmatrix} \begin{matrix} -2 & -2 \\ -2 & 2 \end{matrix} \Bigg|$$

can play the part of that one which appears in any of its first three blocks; this accounts for 12 hyperbolic involutions. As for the remaining four blocks, not only does each provide a hyperbolic involution as it stands but it provides three others—by transposing any two of its three columns and multiplying by -1 either the untransposed column only or all three columns, whichever alternative is the one to restore symmetry to the matrix.

26. It is easy to give the explicit forms for the matrices of the stabiliser Ω_B when B is $y = z = 0$. There are 16 of them; those 8 of Ω_{B+} have $+1$ at their top left-hand corner, zeros elsewhere in the top row and left-hand column, and the residual block one of

$$\begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix} \quad \begin{matrix} -1 & 0 \\ 0 & -1 \end{matrix} \quad \begin{matrix} 0 & 1 \\ -1 & 0 \end{matrix} \quad \begin{matrix} 0 & -1 \\ 1 & 0 \end{matrix} \quad \begin{matrix} 2 & -2 \\ 2 & 2 \end{matrix} \quad \begin{matrix} -2 & 2 \\ -2 & -2 \end{matrix} \quad \begin{matrix} 2 & 2 \\ -2 & 2 \end{matrix} \quad \begin{matrix} -2 & -2 \\ 2 & -2 \end{matrix}.$$

The other 8 matrices have -1 in the top left-hand corner, and the residual blocks are the 8 two-rowed blocks just given but each with its bottom row changed in sign throughout. The 8 matrices with only $0, 1, -1$ for their elements constitute Ω_B^+; here, in contrast to $q = 3, 5$, the subgroups Ω_B^+ differ for different vertices of the same Δ.

The 4 e-points on $x = 0$, being those points for which

$$y = 2z, \quad y = -2z, \quad 2y = z, \quad -2y = z,$$

undergo a cyclic group of 4 permutations under Ω_B^+ when B is $y = z = 0$. Thus $\Omega^+(3, 7)$, transitive on the 4 c-lines through an i-point as well as on the 21 i-points, is transitive on the 28 c-lines, and the stabiliser of a given c-line in $\Omega^+(3, 7)$ is of order 6. Now any c-line is a side of 2 Q, each invariant for an octahedral subgroup of $\Omega^+(3, 7)$ imposing the 4! permutations on its sides; there are 3! operations of this subgroup for which c is invariant and which impose the 3! permutations on the remaining sides. These 3! operations exhaust the stabiliser: any projectivity of $\Omega(3, 7)$ that transposes 2 Q that share a side must lie outside $\Omega^+(3, 7)$. The 14 Q fall, with their diagonal triangles,

into 2 imprimitive systems of 7, and 2 Q with a common side always belong to opposite systems.

27. There is a symmetric (3, 3) correspondence between Δ in opposite systems; two Δ, one of each system, correspond when they share a vertex. There is also a symmetric (4, 4) correspondence between Δ in opposite systems; two Δ, one of each system, correspond when the Q associated with them share a side. This latter is of course the same as the correspondence between Δ of opposite systems that do *not* share a vertex; such Δ are in a certain geometrical relation that will now be obtained.

Let c be one side of a Q; i_1, i_2, i_3 the vertices of Q on c; i'_1, i'_2, i'_3 its opposite vertices. Each pair of opposite vertices is conjugate for χ. The diagonal triangle Δ is $j_1 j_2 j_3$ where j_1 is common to $i_2 i'_3$ and $i'_2 i_3$, and so on. The polars $j_1 i'_1, j_2 i'_2, j_3 i'_3$ of i_1, i_2, i_3 are concurrent at e, the pole of c; the intersection e_1 of c and $j_1 i'_1$ is the point on c conjugate to i_1, and likewise for e_2 and e_3. The s-lines through j_1 join it to the i on $j_2 j_3$ and so meet c at i_1, i_2, i_3, e_1. Thus $j_1 e_2$ and $j_1 e_3$ are c-lines as, likewise, are $j_2 e_3$, $j_2 e_1$, $j_3 e_1$, $j_3 e_2$, and the *only* c-lines through, say, e_1 are $e_1 j_2$, $e_1 j_3$, $i_1 i_2 i_3$.

Take, now, the other Q of which c is a side; it also has i_1, i_2, i_3 for vertices but has another diagonal triangle $k_1k_2k_3$; and the *only* c-lines through e_1 are e_1k_2, e_1k_3, $i_1i_2i_3$. This implies, since k_1, like j_1, is on ee_1, that $j_1j_2j_3$ and $k_1k_2k_3$ are in perspective from e_1. Similarly they are in perspective from e_2 and e_3. And they are manifestly in perspective from e.

Two Δ, in opposite systems and not sharing a vertex, are therefore in quadruple perspective. Their centres of perspective are all e-points and the polar of one of them contains the other three and is the common side of the two Q associated with these Δ.

Just as for $q = 5$, so for $q = 7$; the axes of perspectivity of two Δ can be displayed as factors of linear combinations of the products of their sides. The simplest such identities occur when one Δ is the triangle of reference and the other, to be selected from the lower stratum but not to be any of the first three blocks therein, is the Δ answering to the last block, since the product of its three sides is a symmetric function of the coordinates. The identities, over $GF(7)$, are

$$(2x + 3y + 3z)(3x + 2y + 3z)(3x + 3y + 2z) - xyz \equiv 2(2x - y - z)(2y - z - x)(2z - x - y),$$

$$(2x + 3y + 3z)(3x + 2y + 3z)(3x + 3y + 2z) + xyz \equiv -3(x + y + z)(x^2 + y^2 + z^2 + yz + zx + xy).$$

The centres of perspectivity of these two triangles are then (1, 1, 1) and the three e-points

$$(2, -1, -1), \quad (-1, 2, -1), \quad (-1, -1, 2)$$

on its polar. The Q associated with these Δ are that whose sides are

$$x+y+z=0,\quad y+z-x=0,\quad z+x-y=0,\quad x+y-z=0,$$

and that whose sides are

$$x+y+z=0,\quad y+z+2x=0,\quad z+x+2y=0,\quad x+y+2z=0.$$

Each of the 28 c-lines is a common side of two Q, whose diagonal triangles are in quadruple perspective in the manner described above; there are 28 pairs of identities of which the pair displayed is one, and the other 27 pairs are derivable from this one pair by applying the orthogonal transformations.

28. The symmetrical (3, 3) correspondence between two sets of 7 objects must occur in any representation of $\Omega^+(3, 7)$. For Klein's representation as a group of ternary substitutions over the complex field there occur (**9a**, p. 715; **7**, p. 443) two sets of 7 conics; every conic of either set meets Klein's non-singular plane quartic in the 8 contacts of 4 bitangents, and the 7 sets of 4 bitangents answering to the 7 conics of either set account for all 28 bitangents. Each bitangent belongs to one, and only one, quadruple of either set (**9a**, p. 712) and a symmetrical correspondence between the two sets of 7 conics is set up if conics, one in each set, correspond when the quadruples do not have a bitangent in common.

But perhaps the representation of $\Omega^+(3, 7)$ that most simply displays the (3, 3) correspondence (although the two sets do not now consist of like objects) is the group of 168 projectivities of the 7-point plane $\bar{\omega}$, a point and line corresponding when they are incident. In $\bar{\omega}$ each of the 7 lines contains 3 points and each of the 7 points lies on 3 lines.

29. It was announced by Galois (**6**, p. 412) that $LF(2, q)$ has a permutation representation of degree q for $q = 5, 7, 11$; this is never so if $q > 11$. The isomorphic group $\Omega^+(3, q)$ must therefore also admit such a representation in the finite plane; one has already been encountered for $q = 5, 7$, when the q objects permuted are canonical triangles: for $q = 5$ the whole set, for $q = 7$ the members of either imprimitive system. And so the question is clamant: what geometrical entities supply a representation of $\Omega^+(3, 11)$ as a permutation group of degree 11?

In the finite plane corresponding to $q = 11$ there are, as we shall see, Clebsch hexagons; hexagons, that is, endowed in 10 ways with the Brianchon property of concurrence of 3 diagonals. Given the conic χ there are 22 Clebsch hexagons $\mathscr{C}$ all of whose vertices are e-points and diagonals s-lines; each of the 66 e-points is a vertex of 2 $\mathscr{C}$ that belong one to each of 2 imprimitive systems of 11 $\mathscr{C}$. Either system supplies a representation of $\Omega^+(3, 11)$ as a permutation group of degree 11. The operations of $\Omega(3, 11)$ that are outside $\Omega^+(3, 11)$ transpose the 2 systems.

30. Take χ, $x^2 + y^2 + z^2 = 0$, and the triangle of reference Δ. Suppose that an s-line meets both $y = 0$ and $z = 0$ in points e neither of which is a

vertex of Q; such points are $(c, 0, 1)$ and $(b, 1, 0)$ with b, c both marks of $GF(11)$ and both b^2 and c^2 neither 0 nor 1. Moreover, the points being e, neither $b^2 + 1$ nor $c^2 + 1$ can be a square. Since, in $GF(11)$,

$$\begin{array}{lrrrrrl} & 1, & 4, & -2, & 5, & 3 & \text{are the squares,} \\ \text{and} & -1, & -4, & 2, & -5, & -3 & \text{the non-squares,} \end{array}$$

b^2 and c^2 can only be -2 or 5. But the join

$$x = by + cz$$

is not an s-line unless $b^2 + c^2 + 1$ is a square; this prevents $b^2 + c^2$ from being -4 or -1 and forces it to be 3; b^2 and c^2 are, in either order, the two marks -2 and 5, squares of ± 3 and $\pm\frac{1}{3}$. This yields 2 quadrangles whose 4 vertices are e and 6 joins s, two of the joins being $y = 0$ and $z = 0$; one has vertices $(\pm 3, 1, 0)$ and $(1, 0, \pm 3)$; the other has vertices $(1, \pm 3, 0)$ and $(\pm 3, 0, 1)$.

Consider now the first of these quadrangles. Through each vertex pass, in addition to its joins to the other vertices, 2 further s-lines; the 8 s-lines so arising are found to meet 4 at each of 2 e-points on $x = 0$ and these 8 s-lines, with $x = 0$ and the 6 joins of the quadrangle, are the 15 joins of 6 e-points, namely of

$$30.1 \qquad \begin{array}{rrrrrr} 3 & 1 & 0 & -3 & 1 & 0 \\ 1 & 0 & 3 & 1 & 0 & -3 \\ 0 & 3 & 1 & 0 & -3 & 1. \end{array}$$

Verification is immediate. And the 15 s-lines are sides of the following 5Δ:

$$30.2 \qquad \begin{array}{ll} \Delta_0: & xyz = 0 \\ \Delta_1: & (\ \ 5x - 4y - 2z)(-2x + 5y - 4z)(-4x - 2y + 5z) = 0 \\ \Delta_2: & (\ \ 5x + 4y + 2z)(-2x - 5y + 4z)(-4x + 2y - 5z) = 0 \\ \Delta_3: & (-5x - 4y + 2z)(\ \ 2x + 5y + 4z)(\ \ 4x - 2y - 5z) = 0 \\ \Delta_4: & (-5x + 4y - 2z)(\ \ 2x - 5y - 4z)(\ \ 4x + 2y + 5z) = 0. \end{array}$$

Denote by $\mathscr{C}$ the hexagon whose vertices are 30.1. Each Δ_j in 30.2 answers to a syntheme of vertices of $\mathscr{C}$; the 5 synthemes, one for each Δ_j, constitute a synthematic total T. Each of the 10 synthemes extraneous to T provides 3 pairs whose joins concur: $\mathscr{C}$ has the Clebsch property. The concurrencies are all at points i, and normalized coordinate vectors for them are

$$30.3 \qquad \begin{array}{rrrrrrrrrr} 0 & 5 & 3 & 0 & -5 & 3 & -2 & 2 & 2 & 2 \\ 3 & 0 & 5 & 3 & 0 & -5 & 2 & -2 & 2 & 2 \\ 5 & 3 & 0 & -5 & 3 & 0 & 2 & 2 & -2 & 2. \end{array}$$

Each of these points is the concurrence of sides of 3 of the 5 Δ_j; and each of these 3 sides is an axis of perspective of the 2 remaining Δ_j. The triple perspectivity of Δ_0 and Δ_1 accords with the identity, over $GF(11)$,

$$\begin{aligned} &(5x - 4y - 2z)(-2x + 5y - 4z)(-4x - 2y + 5z) + xyz \\ &\qquad \equiv 4(5x + 4y + 2z)(2x + 5y + 4z)(4x + 2y + 5z). \end{aligned}$$

Other identities derived from this by imposing the orthogonal transformations exhibit other pairs of canonical triangles in triple perspective.

One may, in passing, note the relation between the Q associated with such triangles; they are found to have a common side, the line on which the 3 centres of perspective of their diagonal triangles lie, and their 2 sets of 3 vertices thereon account for all e-points on the side.

31. The elliptic involution centred at a vertex of Δ_0 leaves Δ_0 invariant while transposing the other Δ_j as two pairs; the analogous situation holds for the involution centred at any vertex of any Δ_j, and the Δ_j thus undergo the 15 even permutations of period 2 of the alternating group $\mathfrak{A}_5$. The 15 involutions all belong to $\Omega^+(3, 11)$ and generate a subgroup thereof; this is icosahedral, being isomorphic to $\mathfrak{A}_5$ because any projectivity which imposes the identity permutation of the Δ_j must impose it on the points 30.3 (each of which is determined by those 3 Δ_j whose sides intersect there) and so be the identity projectivity.

$\mathscr{C}$ is not invariant for the whole group $\Omega^+(3, 11)$, it is changed to other hexagons by the involutions centred at the points 30.3; the subgroup for which it is invariant is thus a maximal icosahedral subgroup of order 60, and $\mathscr{C}$ is one of

$$660 \div 60 = 11$$

Clebsch hexagons permuted transitively by $\Omega^+(3, 11)$. The other 10 $\mathscr{C}$ are obtained at once by imposing the involutions centred at the points 30.3; taking, for example, the last of these points $x = y = z$ we have

$$\begin{bmatrix} -4 & -3 & -3 \\ -3 & -4 & -3 \\ -3 & -3 & -4 \end{bmatrix} \begin{bmatrix} 3 & 1 & 0 & -3 & 1 & 0 \\ 1 & 0 & 3 & 1 & 0 & -3 \\ 0 & 3 & 1 & 0 & -3 & 1 \end{bmatrix} = \begin{bmatrix} -4 & -1 & -2 & -2 & -5 & 5 \\ -2 & -4 & -1 & 5 & -2 & -5 \\ -1 & -2 & -4 & -5 & 5 & -2 \end{bmatrix}.$$

The vertices of the 11 $\mathscr{C}$ account for all 66 e-points, and the Δ_j which belong 5 to each $\mathscr{C}$ account for all 55Δ.

32. The 11 $\mathscr{C}$ only provide one half of the figure; there is a second set of 11 Clebsch hexagons $\mathscr{D}$ equally well supplying a permutation representation. These are obtained by starting, instead of from 30.1, from

$$\begin{matrix} 1 & 3 & 0 & 1 & -3 & 0 \\ 3 & 0 & 1 & -3 & 0 & 1 \\ 0 & 1 & 3 & 0 & 1 & -3 \end{matrix}$$

which affords, by a synthematic total of its vertices, the 5 Δ got by transposing y and z throughout 30.2. This transposition is effected by using the involution whose matrix is

$$\begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix};$$

its centre (0, 1, 1) is an e-point so that it is hyperbolic and, although belonging to $\Omega(3, 11)$, it does not belong to $\Omega^+(3, 11)$. Thus $\Omega(3, 11)$ is transitive on 22 Clebsch hexagons $\mathscr{C}$ and $\mathscr{D}$; these form imprimitive systems for $\Omega(3, 11)$ and each set of 11 is a transitive set for $\Omega^+(3, 11)$.

Each e-point is a vertex of a single $\mathscr{C}$ and a single $\mathscr{D}$; there is a symmetrical (6, 6) correspondence between the $\mathscr{C}$ and $\mathscr{D}$ wherein corresponding hexagons share a vertex. Alternatively, one may use the symmetrical (5, 5) correspondence wherein corresponding hexagons do not have a vertex in common. These correspondences between 2 sets of 11 objects will occur in other representations of $\Omega^+(3, 11)$. Klein, in 1879, found a representation as a group of quinary linear substitutions over the complex field, and when the 5 variables on which the substitutions operate are used as homogeneous coordinates in [4] there do occur two sets of 11 quadrics with each quadric of either set linearly dependent on 5 of the other (**9b**, p. 429).

References

1. W. Burnside, *Theory of groups of finite order* (Cambridge, 1911).
2. A. Clebsch, *Über die Anwendung der quadratischen Substitution auf die Gleichung 5ten Grades und die geometrische Theorie des ebenen Fünfseits*, Math. Ann., *4* (1871), 284–345.
3. L. E. Dickson, *Linear groups, with an exposition of the Galois field theory* (Leipzig, 1901).
4. W. L. Edge, *Geometry in three dimensions over GF*(3), Proc. Royal Soc. A*222* (1953), 262–286.
5. ———, *31-point geometry*, Math. Gazette, *39* (1955), 113–121.
6. E. Galois, *Lettre de Galois à M. Auguste Chevalier*, J. de math. pures et appliquées, *11* (1846), 408–415.
7. F. Klein, *Über die Transformation siebenter Ordnung der elliptischen Funktionen*, Math Ann., *14* (1879), 428–471.
8. ———, *Vorlesungen über des Ikosaëder* (Leipzig, 1884).

9a, b. F. Klein and R. Fricke, *Vorlesungen über die Theorie der elliptischen Modulfunktionen* (Leipzig, 1890 and 1892).

10. B. Qvist, *Some remarks concerning curves of the second degree in a finite plane.* Annales Academiae Scientiarum Fennicae A., No. 134 (1952).
11. J. A. Serret, *Cours d'algèbre supérieure* (5th ed.), tome 2 (Paris, 1885).

University of Edinburgh

CONVEX HULLS OF SIMPLE SPACE CURVES

DOUGLAS DERRY

1. Introduction. The convex hull of an arbitrary set M in real Euclidean n-space is known to consist of all the points within the r-simplexes with $r+1$ vertices from M, $r < n$. This note shows that if M is specialized to be a curve A_n of real order n, then its convex hull consists of all the points within the r-simplexes with $r+1$ vertices on A_n, $n = 2r+1$ or $n = 2r$. In the first case each interior point is within exactly one simplex. This result was given by Egerváry **(1)** for $n = 3$. If n is even each interior point of the convex hull of A_n is within a 1-parameter system of $\frac{1}{2}n$-simplexes. The class of curves A_n includes the twisted n-ics, the convex hulls of which have been studied by Karlin and Shapley **(2)**. Some of their results are consequences of the present results.

2. Some definitions. A curve A_n is *defined* to be a 1–1 continuous mapping in real Euclidean n-space of all the real numbers s computed modulo 1 or of the interval, $0 \leqslant s \leqslant 1$, which satisfies the *order condition* that no hyperplane contains more than n points of A_n.

The order condition implies that any linear k-space, $0 \leqslant k < n$, cannot contain more than $k+1$ points of A_n. If a hyperplane H supports A_n at an inner point s' then s' is defined to have multiplicity two within H. By displacing the hyperplanes it is possible to show that the *sharpened order condition*[1] holds that no hyperplane contains more than n points of A_n if each point is counted with its proper multiplicity of one or two.

The symbol $[A, B, \ldots]$ denotes the intersection of all the linear spaces which include the point sets $A, B, \ldots$, while $\{A, B, \ldots\}$ denotes the convex hull of the union of the point sets $A, B, \ldots$. Two sets A and B are said to be separated by a hyperplane H provided A is in one of the closed half spaces bounded by H and B in the other.

3. The boundary of A_n. The following lemma is stated without proof.

LEMMA 1. *If a hyperplane H supports a compact set X, then* $\{H \cap X\} = H \cap \{X\}$.

THEOREM 1. *The boundary of* $\{A_n\}$ *consists of all the points within all the q-simplexes for which the vertices are $q+1$ points of A_n including e endpoints,* $2q \leqslant n - 2 + e$, $(e = 0, 1, 2)$.

Received October 14, 1955.

[1] I should like to thank the referee for the improvements he suggested and in particular for pointing out that the above form of the sharpened order condition, which makes no use of differentiability, was sufficient for the results of this paper.

Proof. If P be a boundary point of $\{A_n\}$, a hyperplane H exists which supports A_n and contains P. Let $s_0, s_1, \ldots, s_q$ be the distinct curve points in $H \cap \{A_n\}$. Because of the order condition, $\{s_0, s_1, \ldots, s_q\}$ is a q-simplex. By Lemma 1,

$$P \in H \cap \{A_n\} = \{H \cap A_n\} = \{s_0, s_1, \ldots, s_q\}.$$

As H supports A_n an interior point s_i of A_n must be included in H twice. Consequently if e denotes the number of endpoints of A_n in H, it follows from the order condition that

$$e + 2(q + 1 - e) < n \text{ or } 2q < n - 2 + e.$$

Thus each boundary point of $\{A_n\}$ is within a q-simplex, with the required properties.

Conversely let P be a point of a q-simplex $\{s_0, s_1, \ldots, s_q\}$ for which $2q < n - 2 + e$. Then a hyperplane exists which contains P and supports A_n. To construct such a hyperplane, for each point s_i interior to A_n, let N_i be an arc $s'_i < s < s_i$ and if A_n is not closed let N', N'' be neighbourhoods of the endpoints 0, 1 respectively. Let H be a hyperplane which contains n points of A_n including all s_i, s'_i and so that the remaining $n - 2(q + 1) + e$ curve points within H are distributed among the arcs N_i, N', N'' in such a way that no arc N_i contains an odd number of these points. This distribution is always possible because if A_n is closed n is even and $e = 0$. If $N_i \to s_i$, $N' \to 0$, $N'' \to 1$ then any limiting position of H contains P and supports A_n. As $P \in \{s_0, s_1, \ldots, s_q\} \subseteq \{A_n\}$, P is a boundary point of $\{A_n\}$. The proof is now complete.

4. The structure of $\{A_n\}$. If $2r = n$ or $2r + 1 = n$, S_r is *defined* to be an r-simplex with interior points of A_n as vertices except for even n when at most one of the vertices may be an endpoint of A_n.

THEOREM 2. *The interior points P of $\{A_n\}$ consist of all the interior points of the simplexes S_r.*

For odd n, S_r is uniquely determined by any one of its interior points P; for even n, S_r is uniquely determined by an interior point P and any one vertex which can be either endpoint of A_n or any arbitrary point of A_n if it is closed.

Proof. We show first that every interior point P of a simplex S_r is an interior point of $\{A_n\}$. As $S_r \subseteq \{A_n\}$ it will be sufficient to show P is not a boundary point of $\{A_n\}$. Let e be the number of vertices of S_r which are endpoints of A_n. If P were a boundary point of $\{A_n\}$ it would be within a hyperplane H which would support $\{A_n\}$. H would also support S_r and consequently, as P is an inner point of S_r, $S_r \subseteq H$. Therefore H would contain $2(r + 1 - e) + e$ points of A_n. This would contradict the order condition as, by the definition of S_r, $e = 0$ if $n = 2r + 1$ and $e < 1$ if $n = 2r$. Hence the inner points of the simplexes S_r are all inner points of $\{A_n\}$.

We next show that a given interior point P of $\{A_n\}$ is an interior point of a simplex S_r. Let a be any real number if A_n is closed and 0 if A_n is open. Denote by $A(a, s')$ the arc of points $s, a < s < s'$. Let s_P be the least upper bound of all s' for which $P \notin \{A(a, s')\}$.

We prove that $P \in \{A(a, s_P)\}$. If this were false, P and $\{A(a, s_P)\}$ would be separated by a hyperplane at a positive distance from $\{A(a, s_P)\}$. This hyperplane would also separate $A(a, s')$ and P for $s' > s_P$ provided s' were sufficiently close to s_P. Consequently $P \notin \{A(a, s')\}$ contrary to the choice of s_P.

P is on a supporting hyperplane of $\{A(a, s_P)\}$. To prove this let s_μ be an increasing sequence which converges to s_P. Because $P \notin \{A(a, s_\mu)\}$ a hyperplane H_μ exists which supports $\{A(a, s_\mu)\}$ and contains P. s_μ can be chosen so that H_μ converges. If H be its limit then $P \in H$ and H supports $\{A(a, s_\mu)\}$. But, as s_μ is arbitrary, H supports $\{A(a, s_P)\}$. From this result, together with the fact that $P \in \{A(a,s_P)\}$, it follows that P is a boundary point of $\{A(a,s_P)\}$.

Consequently, by Theorem 1, a simplex S_q exists which contains P, has vertices on $A(a, s_P)$ and for which $2q < n - 2 + e$, where e is the number of vertices of S_q which are endpoints of $A(a, s_P)$. The vertices of S_q are also on A_n. Let e' be the number of these vertices which are endpoints of A_n. As P is not a boundary point of $\{A_n\}$, $2q > n - 2 + e'$. Therefore $e' < e$ and so $0 < e$. If A_n is open, $e' = e - 1$ as 0 is a common endpoint of A_n and $A(a, s_P)$. The two inequalities yield the result $2q = n - 2 + e$. Hence, if $n = 2r$, then $e = 2$ and $q = r$ and, if $n = 2r + 1$, $e = 1$ and $q = r$. If A_n is closed n is even and $e' = 0$. In this case the inequalities show $e = 2$ and $r = q$. P cannot be a point of a face of S_r for such points, by Theorem 1, are boundary points of $\{A_r\}$. Therefore P is an interior point of the r-simplex S_r which satisfies the requirements of the theorem as $e' = 0$ for odd n and $e' < 1$ for even n. This completes the proof of the first part of the theorem.

For even n, $e = 2$ and consequently a is a vertex of S_r. If A_n is closed a is arbitrary and so in this case, for a given P, an S_r exists with an arbitrary vertex. If A_n is open $a = 0$. After a reversal of orientation of the points on the curve, the other endpoint of A_n can be represented by the number 0. Therefore S_r can be chosen so that either endpoint of A_n is a vertex provided n is even.

Suppose now P is a point within two distinct simplexes with vertices $s_0, s_1, \ldots, s_r; s'_0, s'_1, \ldots, s'_r$ and that P is not in a face of $\{s_0, s_1, \ldots, s_r\}$. Let k, $0 < k < r$, be the number of vertices common to both simplexes. It follows, with the use of the Steinitz replacement theorem, that the space

$$[s_0, s_1, \ldots, s_r, s'_0, s'_1, \ldots, s'_r]$$

has dimension at most $2r - k$. It contains $2(r + 1) - k$ points of A_n. This leads to a contradiction of the order condition unless $2r - k = n$ in which case $k = 0$ and $n = 2r$. This proves, for odd n, that P is within only one simplex S_r and, for even n, that P is never in more than one simplex S_r with a given vertex. The proof is now complete.

Corollary. *Every point P in the interior of $\{A_{2r}\}$ is an interior point of each of two suitably chosen simplexes S_r, S'_r which have no common vertex.*

Proof. If A_{2r} is open each interior point P of $\{A_{2r}\}$ is, by the Theorem, interior to a simplex S_r (S'_r) with the endpoint $s = 0$, $(s = 1)$ as a vertex. If S_r, S'_r were to have a common vertex then, by the Theorem, they would be identical and both endpoints of A_{2r} would be vertices in contradiction to the definition of the simplexes. If A_{2r} is closed the result is clear.

Lemma 2. *If the vertices of two r-simplexes S_r, S'_r which have no common vertex are all on A_{2r} and if an arc of A_{2r} exists which contains two vertices of S_r and no vertex of S'_r, then S_r, S'_r have no point in common.*

Proof. Let $s_0, s_1, \ldots, s_r$, $s_0 < s_1 < \ldots < s_r < s_0 + 1$ $(= s_{r+1})$ be the vertices of S_r. By the hypothesis an arc $s_k \leqslant s \leqslant s_{k+1}$ exists which contains no vertex of S'_r, $0 \leqslant k < r$, if A_{2r} is open and $0 \leqslant k \leqslant r$, if A_{2r} is closed. In the latter case the coordinates may be adjusted so that $0 \leqslant k < r$. As S_r, S'_r have no common vertex, *distinct* curve points $t'_1, t_1, \ldots, t'_r, t_r$ of A_{2r} exist so that

$$t'_1 \leqslant s_0 \leqslant t_1 < t'_2 \leqslant s_1 \leqslant t_2 < \ldots < t'_{k+1} \leqslant s_k < s_{k+1} \leqslant t_{k+1} < \ldots \\ < t'_r \leqslant s_r \leqslant t_r \leqslant t'_1 + 1$$

and so that none of the arcs $t'_1 \leqslant s \leqslant t_i$, $1 \leqslant i \leqslant r$, contains a vertex of S'_r. Let H be the hyperplane $[t'_1, t_1, \ldots, t'_r, t_r]$. As H intersects A_{2r} only in the $2r$ points t'_i, t_i, $1 \leqslant i \leqslant r$, all the points of the arcs $t'_i \leqslant s \leqslant t_i$, $1 \leqslant i \leqslant r$, are either on H or on the same side of H while all the points of A_{2r} not within the above arcs are on the opposite side of H. Thus H separates the vertices of S_r from those of S'_r. Furthermore all the vertices of S'_r are at a positive distance from H. Hence S_r and S'_r have no points in common. The Lemma is now proved.

Convex hulls are defined for affine space. The following result shows that the convex hull $\{A_{2r}\}$ can be defined in terms of projective concepts.

Theorem 3. *If $s_0, s_1, \ldots, s_r$; $s'_0, s'_1, \ldots, s'_r$ are curve points of A_{2r} for which*

$$0 \leqslant s_0 < s'_0 < s_1 < \ldots < s_r < s'_r \leqslant 1,$$

for open A_{2r} and

$$s_0 < s'_0 < s_1 < \ldots < s_r < s'_r < s_0 + 1\ (= s_{r+1})$$

for closed A_{2r} then the interior of $\{A_{2r}\}$ consists of all the intersections

$$[s_0, s_1, \ldots, s_r] \cap [s'_0, s'_1, \ldots, s'_r].$$

Proof. Let P be a given point in the interior of $\{A_{2r}\}$. By the Corollary to Theorem 2, simplexes S_r, S'_r exist, without a common vertex, both of which

contain P as an interior point. Let $s_0, s_1, \ldots, s_r$, $0 \leq s_0 < s_1 < \ldots < s_r < s_0 + 1$ be the vertices of S_r. As S_r, S'_r have the common interior point P it follows from Lemma 2 that each arc $s_i \leq s \leq s_{i+1}$, $0 \leq i < r$, contains exactly one vertex of S'_r. Therefore if $s'_0, s'_1, \ldots, s'_r$ be the vertices of S'_r the subscripts may be adjusted so that, for closed A_{2r},

$$s_0 < s'_0 < s_1 < \ldots < s'_{r-1} < s_r < s'_r < s_0 + 1$$

and, for open A_{2r}, either

$$0 \leq s_0 < s'_0 < s_1 < \ldots < s_r < s'_r \leq 1$$

or

$$0 \leq s'_0 < s_0 < \ldots < s'_r < s_r \leq 1.$$

As P is a common point of the simplexes

$$P \in [s_0, s_1, \ldots, s_r] \cap [s'_0, s'_1, \ldots, s'_r].$$

Now let Q be any point of $[s_0, s_1, \ldots, s_r] \cap [s'_0, s'_1, \ldots, s'_r]$ where

$$s_0, s_1, \ldots, s_r, s'_0, s'_1, \ldots, s'_r$$

are points of A_{2r} which satisfy the inequality system. The r-spaces $[s_0, s_1, \ldots, s_r]$, $[s'_0, s'_1, \ldots, s'_r]$ must have at least one point in common as $2r = n$. They cannot have more than one point in common for then

$$[s_0, s_1, \ldots, s_r, s'_0, \ldots, s'_r]$$

would have dimension at most $2r - 1$ and contain $2r + 2$ points of A_{2r}, in contradiction to the order condition.

Q cannot be a point on a proper face of either simplex $\{s_0, s_1, \ldots, s_r\}$, $\{s'_0, s'_1, \ldots, s'_r\}$. Suppose, for example, Q to be within the face $\{s_0, s_1, \ldots, s_{r-1}\}$. Then the space

$$[s_0, s_1, \ldots, s_{r-1}, s'_0, \ldots, s'_r]$$

would have dimension at most $2r - 1$ and contain $2r + 1$ points of A_{2r} in contradiction to the order condition.

If $s_0, s_1, \ldots, s_r, s'_0, \ldots, s'_r$ move continuously so that the inequalities are always satisfied, Q is uniquely defined and moves continuously. We know, if $Q = P$, that Q is interior to $\{A_{2r}\}$ as well as to both simplexes $\{s_0, s_1, \ldots, s_r\}$, $\{s'_0, s'_1, \ldots, s'_r\}$. As Q cannot enter a proper face of either of these simplexes it must remain in the interior of both of them. Q cannot enter the boundary of $\{A_{2r}\}$. For otherwise it would be in a hyperplane H supporting $\{A_{2r}\}$ and consequently supporting $\{s_0, s_1, \ldots, s_r\}$. As Q is an interior point of the simplex, $[s_0, s_1, \ldots, s_r] \subseteq H$. It follows from the inequality system that at most one vertex of $\{s_0, s_1, \ldots, s_r\}$ is an endpoint of A_{2r}. Hence H would contain at least $2(r + 1) - 1 = 2r + 1$ points of A_{2r} in contradiction to the order condition. Therefore Q must always remain in the interior of $\{A_{2r}\}$. The proof is now complete.

References

1. E. Egerváry, *On the smallest cover of a simple arc of space curve.* Publ. Math. Debrecen *1* (1949), 65–70.
2. S. Karlin and L. S. Shapley, *Geometry of moment spaces.* Mem. Amer. Math. Soc. *12* (1953).

University of British Columbia

CONSTRUCTIONS IN HYPERBOLIC GEOMETRY

FRANS HANDEST

Introduction. In hyperbolic geometry we have three compasses, namely an ordinary compass for drawing ordinary circles with a given centre and a given radius, a hypercompass for drawing hypercycles with a given axis and a given radius, and a horocompass for drawing horocycles with a given diameter and passing through a given point.

Nestorovič has proved that everything that can be constructed by means of one of the compasses and a ruler, can be constructed by means of either of the other compasses and a ruler (**6**; **7**; **8**; **9**). Another important result we want to use in the following is a theorem by Schur concerning ruler constructions. Schur proved that even if we are only able to perform constructions in a finite part Ω' of the projective plane, we are also able to carry out constructions in the entire plane. A point is then said to be constructed if it is determined as the intersection between two lines in Ω'. A line is said to be constructed if there are constructed two points on the line (**11**, pp. 15–22; see also **13**). Another theorem we shall use is: To a given right-angled triangle $\{a, b, c, A, B\}$ (i.e., a right-angled triangle with hypotenuse c, catheti a and b, and opposite angles A and B) there corresponds a second right-angled triangle[1] $\{\Delta(\frac{1}{2}\pi - A), a, \Delta(B), \Pi(c), \frac{1}{2}\pi - \Pi(b)\}$ and using the same transformation on this triangle we obtain a third right-angled triangle and so on. Triangle number six is identical with triangle number one. This sequence of five triangles is called the Engel Chain (**5**, pp. 40–41).

In this paper, we consider the following instruments: parallel-ruler, ruler, compass with fixed adjustment, and hypercompass with fixed adjustment.

1. The parallel-ruler. A parallel-ruler is, as in Euclidean geometry, an instrument for drawing a line through a given point and parallel to a given line. We shall also, as in Euclidean geometry, use the parallel-ruler as an ordinary ruler.

THEOREM 1. *Any construction in hyperbolic geometry that can be performed by means of a ruler and any of the three compasses, can be performed by means of a parallel-ruler.*

Let the hyperbolic plane be the interior of the "absolute" conic Ω situated in the real projective plane. If U and V' are two points[2] on Ω determined by

[1]The angle $\pi(p)$ is the angle of parallelism for the segment of length p. If $A = \pi(p)$ then $p = \Delta(A)$.

[2]In the following, U and $V(U_1, V_1, U', V'$ and so on) will always be points on Ω. If a line intersects Ω at U (or $U_1, U_1', \ldots$) then its other end is called V (or $V_1, V_1', \ldots$), unless otherwise indicated.

the lines u and u', we can always, by means of the parallel-ruler, draw the lines AV' and BU where A and B are two arbitrary points on u and u', respectively, neither of them being the point of intersection $u \cdot u'$. The points U and V' are now determined by pairs of lines. This means, according to the result of Schur, that we are able to join two points on Ω and to perform ruler constructions in the entire projective plane, operating only inside a finite part Ω' of the hyperbolic plane. Of course we have to choose Ω' so that it contains parts of the lines determining the points on Ω. Consequently it is possible to make the following constructions:

1.01. *Given a segment OA on one arm of an angle $V'OV$, construct $OB = OA$ where B is on the other arm of the angle.*

Draw UU' and VV'. Through their intersection draw a line through A. It will meet $U'V'$ at B.

Proof. $UOAV \;\overline{\wedge}\; U'OBV'$ and since $U \rightarrow U'$ and $V \rightarrow V'$, the perspectivity is a congruent transformation that takes OA to OB.

1.02. *Given a segment OA, construct C on the line OA so that $OA = OC$ $(A \neq C)$.*

Draw any line U_1V_1 $(\neq OA)$ through A. Draw U_1O and V_1O and call their other ends (i.e., intersections with Ω) U_1' and V_1', respectively. Then $U_1'V_1'$ intersects OA at C, and $OA = OC$.

1.03 *Given an angle $V'OV$, construct its internal bisector.*

Construct A and B on OV and OV', respectively, so that $OA = OB$ (1.01). AV' and BV will intersect at a point of the angle bisector.

1.04 *Given a line UV and a point P not on the line, construct the perpendicular line to UV through P.*

Bisect the angle UPV (1.03). The angle bisector is perpendicular to UV.

1.05 *Given an angle VOV', construct an angle $VOV'' = 2 \cdot VOV'$.*

Take a point P on OV and construct the symmetric point P' to P with respect to OV' (1.04 and 1.02). Then $VOP' = 2 \cdot VOV'$.

1.06 *Given a line l and a point P on l, construct a line n perpendicular to l through P.*

Draw any ray (not contained in l) beginning at P, and double the angle around l (1.05). Construct the internal bisector n of the supplement of the double angle.

1.07 *Given a segment $AB = p$, construct $\Pi(p)$.*

Construct the perpendicular line l to AB at A (1.06) and draw a parallel line to l through B.

1.08 *Given a segment AB and a point A', both on a line UV, construct B' on UV such that $AB = A'B'$.*

Take any line U_1V_1 ($\neq UV$) through A. If U_2 is the other end of U_1B and V_2 is the other end of V_1A' then B' is $UV \cdot U_2V_2$.

Proof. Use Andrianov's theorem (**1**; for a more elegant proof see **2**): Let the four sides of a quadruply asymptotic crossed quadrangle meet an arbitrary transversal in points A, C, B, D; then BC and DA are congruent segments.

1.09 *Given a segment AB and a ray l' starting at A', construct B' on l', such that $AB = A'B'$.*

Construct B_1 on AA' such that $AB = AB_1$ (1.01) and the point B_2 on AA' such that $AB_1 = A'B_2$ (1.08), and finally the point B' on l' such that $A'B_2 = A'B'$ (1.01).

1.10 *Given a segment AB, construct the mid-point M.*

If A_1A and BB_1 are equal and both perpendicular to AB at A and B, respectively (1.06 and 1.09), and A_1 and B_1 are on opposite sides of AB, then A_1B_1 intersects AB at M.

1.11 *Given two segments a and c ($a < c$), construct a right-angled triangle with hypotenuse c and cathetus a.*

Construct $\Pi(c)$ (1.07) and use $\Pi(c)$ and a to construct the second triangle of the Engel chain (cathetus a fixed) (1.09 and 1.06), so as to obtain B as the angle of parallelism of the hypotenuse (1.07). Construct then the right-angled triangle containing this angle B and the adjacent cathetus a (1.09 and 1.06) The hypotenuse is c and the required triangle is constructed.

1.12 *Given a point O, a segment $r = AB$ and a line l intersecting the circle $O(r)$, construct the points of intersection.*

Construct OO_1 perpendicular to l (with O_1 on l) (1.04) and the right-angled triangle with hypotenuse AB and cathetus OO_1 (1.11). The other cathetus O_1C can now be moved to l (1.09). C_1 and C_2 (where $O_1C = O_1C_1 = O_1C_2$) are the intersections.

1.13 *Given two points O_1 and O_2 and two segments of length r_1 and r_2, construct the intersections of the circles $O_1(r_1)$ and $O_2(r_2)$.*

Let d denote the distance O_1O_2, and b the distance from O_1 to the intersection of O_1O_2 and the radical axis; then

$$\tanh b = \frac{\cosh r_1 \cosh d - \cosh r_2}{\cosh r_1 \sinh d}.$$

The segment b can be constructed in the following way: Construct $O_1O_1' = r_1$ and $O_2O_2' = r_2$, both perpendicular to O_1O_2 (1.06 and 1.09), with O_1' and O_2'

on the same side of O_1O_2. Construct the mid-point M of $O_1'O_2'$ (1.10) and construct the line m perpendicular to $O_1'O_2'$ at M (1.06). Let m meet O_1O_2 at A; then $AO_2 = b$.

Proof. $O_1'A = AO_2'$. If $AO_2 = x$, so that $O_1A = d - x$, then

$$\tanh x = \frac{\cosh r_1 \cosh d - \cosh r_2}{\cosh r_1 \sinh d}.$$

Since tanh is a single-valued function, we have $x = b$ and 1.13 reduces to 1.12 (**8**).

Any construction that can be performed by means of a compass and a ruler can then be performed by means of a parallel-ruler, and this result, along with the theorem of Nestorovič, proves Theorem 1.

2. Analogues of Steiner's construction

THEOREM 2. *Any construction that can be performed by means of any of the three compasses and ruler, can be carried out with the ruler alone if there is drawn somewhere in the plane* (i) *a circle with its centre and two parallel lines, or* (ii) *a hypercycle with its axis and two parallel lines with their common end not on the axis, or* (iii) *a horocycle with one diameter and two parallel lines with their common end not at the centre of the horocycle* (**12**).

(i) Let Ω again be the absolute conic, ω the given circle with centre A, and P the common end of the two given parallel lines. We want to prove that if O is any given ordinary point and l is any given line, we are able to construct the parallels from O to l. When this is proved, Theorem 1 will give us Theorem 2(i). Let Ω' be a finite part of the hyperbolic plane containing ω, O, a part of l, and a part of the two lines that define P. By means of two harmonic constructions, we can obtain the polar a of A with respect to ω. This is also the absolute polar of A (i.e., the polar with respect to Ω). The construction can be carried out by using the ruler only inside Ω'. Join P and A and let Q be one of its intersections with ω. The homology H, with axis a, centre A, taking Q to P, will take ω to Ω (**4**, pp. 173–174). H^{-1} will then take Ω to ω.

Construct now the images O' and l' of O and l in the homology H^{-1}. Join O' to the intersections, P_1 and P_2, of ω and l', and construct the images of $O'P_1$ and $O'P_2$ in the homology H. These lines are the parallel lines desired.

(ii). Given a hypercycle ω, with axis a, and two parallel lines with end P (P not on a), we can again use two harmonic constructions to obtain the pole A of a with respect to ω. The constructions can be carried out by using the ruler inside a suitable finite part Ω' of the hyperbolic plane. The point A is also the absolute pole of a. Let AP intersect ω at Q, as before. The homology H, with axis a, centre A, taking Q to P, will take ω to Ω. Using the same principle as above, we are able to construct a line through a given point parallel to a given line. This proves Theorem 2(ii).

(iii). In the third case, where we have a horocycle ω with centre A, the homology is an elation. But here, the centre A is not a given point. To determine A, we have to construct a second diameter of the horocycle. This can be done as follows: Let B be the ordinary end-point of the given diameter d, and let F be any other given point on ω (neither B nor A). Choose on the conic three distinct points C, D, E, none of them coincident with B or F, and let l be the join of the intersections $d \cdot DE$ and $BC \cdot EF$. Join F to the intersection $l \cdot CD$. Since this is a line passing through A, it is a diameter. For, l is the Pascal line of the hexagon $ABCDEF$.

The centre A is now determined by two parallel lines. The tangent a to ω can be constructed as the Pascal line of the hexagon $AABCDE$. This is also the tangent at A to Ω. All the above constructions can be performed by ruler inside a suitable finite part Ω' of the hyperbolic plane. The elation, with centre A and axis a, taking Q to P (where P is the given end and Q an intersection of AP and ω), plays now the same role as the homology H in (i) and (ii).

As shown by Obláth in connection with Steiner constructions (**10**; for a more elegant proof see **3**), it is sufficient if we are given only an arc, however small, of the circle, hypercycle, or horocycle. Hüttemann's proof, being projective, is valid here.

3. Compasses with fixed adjustment

THEOREM 3. *Every construction that can be performed by any one of the three compasses and ruler can be performed by either* (i) *a compass with fixed adjustment and a ruler or* (ii) *a hypercompass with fixed adjustment and a ruler.*

If we can prove that by means of our instruments we are able to construct a pair of parallel lines, then Theorem 3 will follow from Theorem 2.

(i) Draw a circle ω with centre A and a diameter l. Construct (by means of two harmonic constructions) the pole L of l with respect to ω. This is also the absolute pole of l. Given a point P either on l or outside l, PL is then perpendicular to l. All the constructions can be performed inside a suitable part Ω' of the hyperbolic plane.

The usual parallel construction can now be carried out, taking the arbitrary radius to be the radius given by the adjustment.

(ii). Perpendicular lines can be constructed in the same way as in (i), using a hypercycle instead of a circle.

Two parallel lines can be constructed in the following way: Draw an acute angle AOB and construct on OA a point A_1 so that OA_1 is equal to the adjustment of the hypercompass. Construct l perpendicular to OB at O and l_1 perpendicular to l through A_1. Let the hypercycle with axis OB intersect l_1 at S. The line m perpendicular to OB through S is parallel to OA. As a matter of fact, this is only the usual parallel construction here performed by a hypercompass instead of the ordinary compass.

4. The common perpendicular to two skew lines. Finally we wish construct (e.g., by means of ruler and compass) the common perpendicular to two skew lines in hyperbolic 3-space.

Let the given lines be g and g'. Take an arbitrary point A on g and construct the two lines AE_1 and AE_2, where E_1 and E_2 are the ends of g'. On g, construct points M_1, M_2, such that M_iE_i is parallel to AE_i and perpendicular to $g(i = 1 \text{ or } 2)$. If M is the mid-point of M_1M_2 and MN is perpendicular to g', then MN is the required common perpendicular.

Proof. Project the whole figure on the plane Ng. If the projections of E_1 and E_2 are F_1 and F_2, respectively, then $MM_1F_1N \equiv MM_2F_2N$ and therefore $\angle M_1MN = \angle M_2MN = \frac{1}{2}\pi$.

References

1. S. N. Andrianov, *A synthetic demonstration of a theorem in Lobachevskian geometry* (Russian), Učenye Zapiski Kazan, Univ. *101*, kn 3 (1941), 22–23.
2. H. S. M. Coxeter, *Review of Andrianov's paper*, Math. Rev., *10* (1949), 319.
3. F. Hüttemann, *Ein Beitrag zu den Steinerschen Konstruktionen*, Jber. dtsch. Math. Ver., *43* (1934), 184–185.
4. C. Juel, *Vorlesungen über projektive Geometrie* (Berlin, 1934).
5. H. Liebmann, *Nichteuklidische Geometrie*, Sammlung Schubert XLIX (Leipzig, 1912).
6. N. M. Nestorovič, *Sur l'équivalence par rapport à la construction du complexe MB et du complexe E*, C.R. Acad. Sci. URSS (N.S.), *22* (1939), 224–227.
7. ———, *Sur la puissance constructive d'un complexe E sur le plan de Lobatchevski*, C.R. Acad. Sci. URSS (N.S.), *43* (1944), 186–188.
8. ———, *Geometrical constructions with horocycle-compass and ruler in the Lobačevskiĭ plane* (Russian), Doklady Akad. Nauk. SSSR (N.S.), *66* (1949), 1047–1050.
9. ———, *On the equivalence of a hypercircle to an ordinary circle in constructions in the Lobačevskiĭ plane* (Russian), Doklady Akad. Nauk. SSSR (N.S.), *69* (1949), 731–735.
10. R. Obláth, *Bemerkungen zur Theorie der geometrischen Konstruktionen*, Monatsh. Math. Phys. *26* (1915), 295–298.
11. F. Schur, *Grundlagen der Geometrie* (Leipzig, 1909).
12. A. S. Smogorževskiĭ, *Geometric constructions in the Lobačevskiĭ plane* (Serbo-croatian), Gosudarstv. Izdat. Tehn.-Teor. Lit., (Moscow and Leningrad, 1951).
13. P. Zühlke, *Konstruktionen in begrenzter Ebene*, Mathematische Bibliothek XI (Leipzig, 1913).

University of Toronto

LOCAL CONNECTEDNESS OF EXTENSION SPACES

BERNHARD BANASCHEWSKI

1. Introduction. An extension E^* of a topological space E (that is, a space containing E as a dense subspace)[1] determines a family of filters $\mathfrak{S}(u)$ on E, given by the traces $U \cap E$ of the neighbourhoods $U \subseteq E^*$ of each $u \in E^* - E$. Many topological properties of an extension E^* of a given space E can be related to properties of these trace filters (as we shall call them) belonging to E^*. In this respect, the following condition for filters $\mathfrak{A}$ has proved to be of some interest:

(C) If $O \cup P \in \mathfrak{A}$, O and P disjoint open sets, then either $O \in \mathfrak{A}$ or $P \in \mathfrak{A}$.

If, for instance, the trace filters of a locally connected extension E^* of a simply connected space E fulfil (C), then E^* is also simply connected **(2)**. This statement involves previous knowledge of the local connectedness of E^*. In the present note, a simple characterisation in terms of trace filters will be given for the local connectedness of extension spaces whose trace filters satisfy condition (C). This will then enable us to show that certain types of extensions, amongst them the Čech compactification of locally compact spaces which are denumerable at infinity, can never be locally connected.

2. The principal result. A filter $\mathfrak{A}$ on a topological space E is called *open*, if it has a basis consisting of open sets. Open filters for which condition (C) holds we shall call *connected*. As one can readily see, condition (C) for open filters is an extension of the concept of connectedness from open sets to open filters. If E^* be an extension of E, to any open set $O \subseteq E$ let $\tilde{O}$ be the set of all points $u \in E^* - E$ which satisfy $O \in \mathfrak{S}(u)$, $\mathfrak{S}(u)$ being the trace filter belonging to u. *$O \cup \tilde{O}$ is open in E^*.* There is at least one open $O^* \subseteq E^*$ for which $O = O^* \cap E$. Obviously, one has $O^* \subseteq O \cup \tilde{O}$, hence it follows that each $x \in O$ is an interior point of $O \cup \tilde{O}$ (in E^*). Furthermore, to each $u \in \tilde{O}$ there exists, by definition, an open neighbourhood V in E^* for which $V \cap E \subseteq O$, and again one has $V \subseteq O \cup \tilde{O}$.

The passage from O to $O \cup \tilde{O}$ will be used as the main device in proving the following proposition:

Let E^* be an extension of E each of whose trace filters is connected. Then E^* is locally connected if and only if E is locally connected and each trace filter has a basis consisting of connected open sets.

Received January 4, 1956.

[1]All topological concepts are used in the sense of Bourbaki **(3)**. All spaces considered here are assumed to be separated (=Hausdorff) spaces.

If E is locally connected and the trace filters $\mathfrak{S}(u)$ have the stated property, it is quite obvious that E^* is locally connected: To each neighbourhood W of $u \in E^*$, there is a connected open set (in E) $V \subseteq W \cap E$ in $\mathfrak{S}(u)$ or in the neighbourhood filter $\mathfrak{B}(u)$ of u in E if $u \in E$, and therefore an open neighbourhood U of u in E^* for which $U \cap E = V$ and $U \subseteq W$. As V is dense in U, U is also connected.

Now, let E^* be locally connected and U any open set (in E) from $\mathfrak{S}(u)$. Then there exists a connected open neighbourhood (in E^*) $V \subseteq U \cup \tilde{U}$ of u (as $U \cup \tilde{U}$ is open, hence a neighbourhood of u) and $W = V \cap E \subseteq U$ belongs to $\mathfrak{S}(u)$. The set $W \cup \tilde{W}$ contains V and, apart from that, only adherence points of V; therefore, the connectedness of V implies that of $W \cup \tilde{W}$. Supposing there were a decomposition $W = O \cup P$ of W into disjoint open sets O and P in E. Then it would follow that $\tilde{W} = \tilde{O} \cup \tilde{P}$, because, by hypothesis, $O \cup P \in \mathfrak{S}(v)$ implies $O \in \mathfrak{S}(v)$ or $P \in \mathfrak{S}(v)$ for any $v \in E^* - E$. This would, however, mean that $(O \cup \tilde{O}) \cup (P \cup \tilde{P})$ is a decomposition of $W \cup \tilde{W}$ into open disjoint sets, in contradiction to the connectedness of $W \cup \tilde{W}$. Conseqently, W is a connected set, and as $W \subseteq U$ and U was arbitrary, this shows $\mathfrak{S}(u)$ has a basis consisting of connected open sets. The same argument applied to the neighbourhood filter $\mathfrak{B}(x)$ of each $x \in E$ (instead of the $\mathfrak{S}(u)$) proves that each $\mathfrak{B}(x)$ also has a basis consisting of connected open sets, or, in other words, that E is locally connected.

As we have proved recently **(2)** the maximal open, the maximal regular and the maximal completely regular filters of a space E are connected filters. It is well known that the non-convergent filters in each of these categories form the set of trace filters of certain extensions of E: the maximal open filters correspond **(4)** to Katětov's maximal Hausdorff extension κE of E; the maximal regular filters, in the case of a regular E, to Alexandroff's **(1)** extension $\alpha' E$; the maximal completely regular filters, in the case of a completely regular E, to Čech's **(1; 2**, ch. IX, 1, ex. 7**)** compactification βE of E. As a corollary to the above proposition, we therefore have:

If E is not locally connected, then κE, $\alpha' E$, and βE are not locally connected either.

3. Application to particular types of spaces. We now want to prove a similar statement for βE, $\alpha' E$ and κE in the case of certain types of spaces E which include locally connected spaces as well as others.

Let E be completely regular and suppose there exist denumerably many open sets $O_i \subseteq E$ whose closures are mutually disjoint and have a closed union. *Then βE is not locally connected.* In each O_i one can find a descending sequence $O_{i,k}(k = 0, 1, 2, \ldots ; O_{i,0} = O_i)$ of open sets such that for each pair $O_{i,k}$, $O_{i,k+1}$ there exists a continuous function $h_{i,k}$ on E, for which

$$0 \leq h_{i,k}(x) \leq 1, \; x \in E;$$

$h_{i,k}(x) = 1$ on $O_{i,k+1}$ and $h_{i,k}(x) = 0$ outside $O_{i,k}$. Now the sets

$$M_i = \bigcup_{s>i} O_{s,i}$$

constitute the basis of a completely regular filter $\mathfrak{N}$: one has $M_{i+1} \subseteq M_i$. Furthermore, the function

$$h_i(x) = \begin{cases} h_{s,i}(x), & x \in O_{s,i},\ s > i \\ 0 & \text{otherwise} \end{cases}$$

is continuous, since

$$\bar{M}_i = \bigcup_{s>i} \bar{O}_{s,i}$$

(by hypothesis concerning the O_i) vanishes outside M_i, is equal to 1 on M_{i+1}, and assumes only values between 0 and 1. This shows the filter $\mathfrak{N}$ is completely regular.

Now, if βE were locally connected, each of the corresponding trace filters (that is, each maximal completely regular filter) would have a basis of connected open sets as proved above. Let $\mathfrak{M} \supseteq \mathfrak{N}$ be maximal completely regular. By Zorn's lemma, or similarly, owing to the compactness of βE, there exist such $\mathfrak{M}$. Then, as $M_0 \in \mathfrak{M}$, there would exist a connected open set $G \subseteq M_0$ in $\mathfrak{M}$, and as M_0 is the union of the disjoint open sets O_i, one would have $G \subseteq O_r$ for a certain r. This, however, would entail $G \cap M_{r+1} = \phi$, in contradiction to $\mathfrak{N} \subseteq \mathfrak{M}$, which proves βE is not locally connected.

In exactly the same way, one obtains the following similar proposition: *If a regular space E contains denumerably many open sets O_i whose closures are mutually disjoint and have a closed union, then the extension $\alpha' E$ of E is not locally connected.* Here one has to construct a regular filter $\mathfrak{N}$ from sequences $O_{i,k}$, $O_{i,0} = O_i$, for which $\overline{O_{i,k+1}} \subseteq O_{i,k}$ holds, and then the proof proceeds as above.

Finally, the same method gives this result: *If a space E contains denumerably many disjoint open sets O_i, then the extension κE is not locally connected.* In this case, one need only consider the open filter generated by the sets

$$\bigcup_{i \geq n} O_i, \quad (n = 0, 1, 2, \ldots)$$

instead of the filters $\mathfrak{N}$ above.

A class of spaces satisfying the hypothesis required for E in the preceding arguments are the locally compact spaces which are denumerable at infinity. A space E of this type is the union of an ascending sequence M_i ($i = 0, 1, 2, \ldots$) of open, relatively compact sets for which $\overline{M_i} \subseteq M_{i+1}$. Then $M_{i+1} - \overline{M_i}$ are disjoint open sets and any collection of open sets O_i satisfying $\overline{O_i} \subseteq M_{i+1} - \overline{M_i}$ will have the desired property: for any $x \in \overline{\bigcup O_i}$ one has $x \in M_s$ for some s and hence $x \in \overline{O_s}$. This means $\bigcup \overline{O_i}$ is closed.

We have, therefore, the following corollary: *For locally compact spaces E denumerable at infinity, none of the extensions βE, $\alpha' E$ and κE is locally connected.*

REFERENCES

1. P. Alexandroff, *Bikompakte Erweiterung topologischer Räume*, Mat. Sbornik, N.S., *5* (1939), 420–428.
2. B. Banaschewski, *Ueberlagerungen von Erweiterungsräumen*, to appear in Archiv der Mathematik.
3. N. Bourbaki, *Topologie générale* (Act. sci. industr., Paris).
4. M. Katětov, *Über H-abgeschlossene und bikompakte Räume*, Časopis Mat. Fys., *69* (1939–40), 36–49.

Hamilton College, McMaster University

MAXIMAL FLOW THROUGH A NETWORK

L. R. FORD, JR. AND D. R. FULKERSON

Introduction. The problem discussed in this paper was formulated by T. Harris as follows:

"Consider a rail network connecting two cities by way of a number of intermediate cities, where each link of the network has a number assigned to it representing its capacity. Assuming a steady state condition, find a maximal flow from one given city to the other."

While this can be set up as a linear programming problem with as many equations as there are cities in the network, and hence can be solved by the simplex method **(1)**, it turns out that in the cases of most practical interest, where the network is planar in a certain restricted sense, a much simpler and more efficient hand computing procedure can be described.

In §1 we prove the minimal cut theorem, which establishes that an obvious upper bound for flows over an arbitrary network can always be achieved. The proof is non-constructive. However, by specializing the network (§2), we obtain as a consequence of the minimal cut theorem an effective computational scheme. Finally, we observe in §3 the duality between the capacity problem and that of finding the shortest path, via a network, between two given points.

1. The minimal cut theorem. A *graph* G is a finite, 1-dimensional complex, composed of *vertices* $a, b, c, \ldots, e$, and *arcs* $\alpha(ab), \beta(ac), \ldots, \delta(ce)$. An arc $\alpha(ab)$ *joins* its end vertices a, b; it passes through no other vertices of G and intersects other arcs only in vertices. A *chain* is a set of distinct arcs of G which can be arranged as $\alpha(ab)$, $\beta(bc)$, $\gamma(cd), \ldots, \delta(gh)$, where the vertices $a, b, c, \ldots, h$ are distinct, i.e., a chain does not intersect itself; a chain *joins* its end vertices a and h.

We distinguish two vertices of G: a, the *source*, and b, the *sink*.[1] A *chain flow* from a to b is a couple $(C; k)$ composed of a chain C joining a and b, and a non-negative number k representing the flow along C from source to sink.

Each arc in G has associated with it a positive number called its *capacity*. We call the graph G, together with the capacities of its individual arcs, a *network*. A *flow* in a network is a collection of chain flows which has the property that the sum of the numbers of all chain flows that contain any arc is no greater than the capacity of that arc. If equality holds, we say the arc is *saturated* by the flow. A chain is *saturated* with respect to a flow if it contains

Received September 20, 1955.

[1]The case in which there are many sources and sinks with shipment permitted from any source to any sink is obviously reducible to this.

a saturated arc. The value of a flow is the sum of the numbers of all the chain flows which compose it.

It is clear that the above definition of flow is not broad enough to include everything that one intuitively wishes to think of as a flow, for example, sending trains out a dead end and back or around a circuit, but as far as effective transportation is concerned, the definition given suffices.

A *disconnecting set* is a collection of arcs which has the property that every chain joining a and b meets the collection. A disconnecting set, no proper subset of which is disconnecting, is a *cut*. The *value* of a disconnecting set D (written $v(D)$) is the sum of the capacities of its individual members. Thus a disconnecting set of minimal value is automatically a cut.

THEOREM 1. (Minimal cut theorem). *The maximal flow value obtainable in a network N is the minimum of $v(D)$ taken over all disconnecting sets D.*

Proof. There are only finitely many chains joining a and b, say n of them. If we associate with each one a coordinate in n-space, then a flow can be represented by a point whose jth coordinate is the number attached to the chain flow along the jth chain. With this representation, the class of all flows is a closed, convex polytope in n-space, and the value of a flow is a linear functional on this polytope. Hence, there is a maximal flow, and the set of all maximal flows is convex.

Now let S be the class of all arcs which are saturated in every maximal flow.

LEMMA 1. *S is a disconnecting set.*

Suppose not. Then there exists a chain $\alpha_1, \alpha_2, \ldots, \alpha_m$ joining a and b with $\alpha_i \notin S$ for each i. Hence, corresponding to each α_i, there is a maximal flow f_i in which α_i is unsaturated. But the average of these flows,

$$f = \frac{1}{m} \sum f_i,$$

is maximal and α_i is unsaturated by f for each i. Thus the value of f may be increased by imposing a larger chain flow on $\alpha_1, \alpha_2, \ldots, \alpha_m$, contradicting maximality.

Notice that the orientation assigned to an arc of S by a positive chain flow of a maximal flow is the same for all such chain flows. For suppose first that (C_1, k_1), (C_2, k_2) are two chain flows occurring in a maximal flow f, $k_1 \geqslant k_2 > 0$, where

$$C_1 = \alpha_1(a\, a_1),\ \alpha_2(a_1 a_2), \ldots, \alpha_j(a_{j-1}, a_j), \ldots, \alpha_r(a_{r-1}, b)$$

$$C_2 = \beta_1(a\, b_1),\ \beta_2(b_1 b_2), \ldots, \beta_k(b_{k-1}, b_k), \ldots, \beta_s(b_{s-1}, b),$$

and $\alpha_j(a_{j-1}, a_j) = \beta_k(b_{k-1}, b_k) \in S$, $a_{j-1} = b_k$, $a_j = b_{k-1}$. Then

$$C_1' = \alpha_1, \alpha_2, \ldots, \alpha_{j-1}, \beta_{k+1}, \ldots, \beta_s$$

$$C_2' = \beta_1, \beta_2, \ldots, \beta_{k-1}, \alpha_{j+1}, \ldots, \alpha_r$$

contain chains C_1'', C_2'' joining a and b, and another maximal flow can be obtained from f as follows. Reduce the C_1 and C_2 components of f each by k_2, and increase each of the C_1'' and C_2'' components by k_2. This unsaturates the arc α_j, contradicting its definition as an element of S. On the other hand, if (C_1, k_1), (C_2, k_2) were members of distinct maximal flows f_1, f_2, consideration of $f = \frac{1}{2}(f_1 + f_2)$ brings us back to the former case. Hence, the arcs of S have a definite orientation assigned to them by maximal flows. We refer to that vertex of an arc $\alpha \in S$ which occurs first in a positive chain flow of a maximal flow as the *left* vertex of α.

Now define a *left arc* of S as follows: an arc α of S is a left arc if and only if there is a maximal flow f and a chain $\alpha_1, \alpha_2, \ldots, \alpha_k$ (possibly null) joining a and the left vertex of α with no α_i saturated by f. Let L be the set of left arcs of S.

LEMMA 2. *L is a disconnecting set.*

Given an arbitrary chain $\alpha_1(a\, a_1), \alpha_2(a_1 a_2), \ldots, \alpha_m(a_{m-1}\, b)$ joining a and b, it must intersect S by Lemma 1. Let $\alpha_t(a_{t-1}, a_t)$ be the first $\alpha_i \in S$. Then for each α_i, $i < t$, there is a maximal flow f_i in which α_i is unsaturated. The average of these flows provides a maximal flow f in which $\alpha_1, \alpha_2, \ldots, \alpha_{t-1}$ are unsaturated. It remains to show that this chain joins a to the left vertex of α_t, i.e., a_{t-1} is the left vertex of α_t. Suppose not. Then the maximal flow f contains a chain flow

$$[\beta_1(ab_1),\ \beta_2(b_1, b_2), \ldots, \beta_r(b_{r-1}, b); k],\quad k > 0,\quad \beta_s = \alpha_t,\quad b_{s-1} = a_t,\quad b_s = a_{t-1}.$$

Let the amount of unsaturation in f of α_i $(i = 1, \ldots, t-1)$, be $k_i > 0$. Now alter f as follows: decrease the flow along the chain $\beta_1, \beta_2, \ldots, \beta_r$ by $\min [k, k_i] > 0$ and increase the flow along the chain contained in

$$\alpha_1, \alpha_2, \ldots, \alpha_{t-1}, \beta_{s+1}, \ldots, \beta_r$$

by this amount. The result is a maximal flow in which α_t is unsaturated, a contradiction. Hence $\alpha_t \in L$.

LEMMA 3. *No positive chain flow of a maximal flow can contain more than one arc of L.*

Assume the contrary, that is, there is a maximal flow f_1 containing a chain flow

$$[\beta_1(ab_1),\ \beta_2(b_1 b_2), \ldots, \beta_r(b_{r-1}, b); k],\ k > 0,$$

with arcs $\beta_i, \beta_j \in L$, β_i occurring before β_j, say, in the chain. Let f_2 be that maximal flow for which there is an unsaturated chain

$$\alpha_1(aa_1),\ \alpha_2(a_1, a_2), \ldots, \alpha_s(a_{s-1}, b_{j-1})$$

from a to the left vertex of β_j. Consider $f = \frac{1}{2}(f_1 + f_2)$. This maximal flow contains the chain flow $[\beta_1, \beta_2, \ldots, \beta_r; k']$ with $k' \geq \frac{1}{2}k$, and each $\alpha_i (i = 1, \ldots, s)$ is unsaturated by $k_i > 0$ in f. Again alter f: decrease the flow along

$\beta_1, \beta_2, \ldots, \beta_r$ by min $[k', k_i] > 0$ and increase the flow along the chain contained in $\alpha_1, \alpha_2, \ldots, \alpha_s, \beta_j, \ldots, \beta_r$ by the same amount, obtaining a maximal flow in which β_i is unsaturated, a contradiction.

Now to prove the theorem it suffices only to remark that the value of every flow is no greater than $v(D)$ where D is any disconnecting set; and on the other hand we see from Lemma 3 and the definition of S that in adding the capacities of arcs of L we have counted each chain flow of a maximal flow just once. Since by Lemma 2 L is a disconnecting set, we have the reverse inequality. Thus L is a minimal cut and the value of a maximal flow is $v(L)$.

We shall refer to the value of a maximal flow through a network N as the *capacity* of N (cap (N)). Then note the following corollary of the minimal cut theorem.

COROLLARY. *Let A be a collection of arcs of a network N which meets each cut of N in just one arc. If N' is a network obtained from N by adding k to the capacity of each arc of A, then* cap $(N') =$ cap $(N) + k$.

It is worth pointing out that the minimal cut theorem is not true for networks with several sources and corresponding sinks, where shipment is restricted to be from a source to its sink. For example, in the network (Fig. 1) with shipment from a_i to b_i and capacities as indicated, the value of a minimal disconnecting set (i.e., a set of arcs meeting all chains joining sources and corresponding sinks) is 4, but the value of a maximal flow is 3.

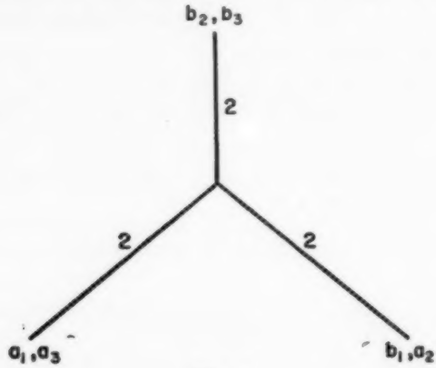

Fig. 1

2. A computing procedure for source-sink planar networks.[2] We say that a network N is *planar* with respect to its source and sink, or briefly, N is *ab*-planar, provided the graph G of N, together with arc ab, is a planar

[2]It was conjectured by G. Dantzig, before a proof of the minimal cut theorem was obtained, that the computing procedure described in this section would lead to a maximal flow for planar networks.

graph (**2**; **3**). (For convenience, we suppose there is no arc in G joining a and b.) The importance of ab-planar networks lies in the following theorem.

THEOREM 2. *If N is ab-planar, there exists a chain joining a and b which meets each cut of N precisely once.*

Proof. We may assume, without loss of generality, that the arc ab is part of the boundary of the outside region, and that G lies in a vertical strip with a located on the left bounding line of the strip, b on the right. Let T be the chain joining a and b which is top-most in N. T has the desired property, as we now show. Suppose not. Then there is a cut D, at least two arcs of which are in T. Let these be α_1 and α_2, with α_1 occurring before α_2 in following T from a to b. Since D is a cut, there is a chain C_1 joining a and b which meets D in α_1 only. Similarly there is a chain C_2 meeting D in α_2 only. Let C_2' be that part of C_2 joining a to an end point of α_2. It follows from the definition of T that C_1 and C_2' must intersect. But now, starting at a, follow C_2' to its last intersection with C_1, then C_1 to b. We thus have a chain from a to b not meeting D, contradicting the fact that D is a cut.

Symmetrically, of course, the bottom-most chain of N has the same property.

Notice that this theorem is not valid for networks which are not ab-planar. A simple example showing this is provided by the "gas, water, electricity" graph (Fig. 2), in which every chain joining a and b meets some cut in three arcs.

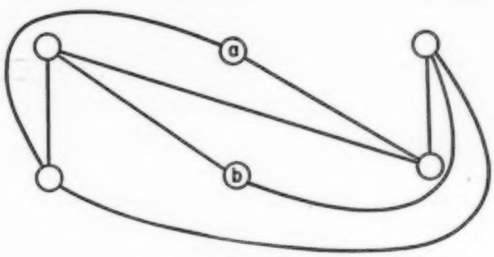

Fig. 2

Theorem 2 and the corollary to Theorem 1 provide an easy computational procedure for determining a maximal flow in a network of the kind here considered. Simply locate a chain having the property of Theorem 2; this can be done at a glance by finding the two regions separated by arc ab, and taking the rest of the boundary of either region (throwing out portions of the boundary where it has looped back and intersected itself, so as to get a chain). Impose as large a chain flow $(T; k)$ as possible on this chain, thereby saturating one or more of its arcs. By the corollary, subtracting k from each capacity in T reduces the capacity of N by k. Delete the saturated arcs, and proceed as

before. Eventually, the graph disconnects, and a maximal flow has been constructed.

3. A minimal path problem. For source-sink planar networks, there is an interesting duality between the problem of finding a chain of minimal capacity-sum joining source and sink and the network capacity problem, which lies in the fact that chains of N joining source and sink correspond to cuts (relative to two particular vertices) of the dual[3] of N and vice versa. More precisely, suppose one has a network N, planar relative to two vertices a and b, and wishes to find a chain joining a and b such that the sum of the numbers assigned to the arcs of the chain is minimal. An easy way to solve this problem is as follows. Add the arc ab, and construct the dual of the resulting graph G. Let a' and b' be the vertices of the dual which lie in the regions of G separated by ab. Assign each number of the original network to the corresponding arc in the dual. Then solve the capacity problem relative to a' and b' for the dual network by the procedure of §2. A minimal cut thus constructed corresponds to a minimal chain in the original network.

---

[3]The *dual* of a planar graph G is formed by taking a vertex inside each region of G and connecting vertices which lie in adjacent regions by arcs. See (**2**; **3**).

References

1. G. B. Dantzig, *Maximization of a linear function of variables subject to linear inequalities: Activity analysis of production and allocation* (Cowles Commission, 1951).

2. H. Whitney, *Non-separable and planar graphs*, Trans. Amer. Math. Soc., *34* (1932), 339–362.

3. ———, *Planar graphs*, Fundamenta Mathematicae, *21* (1933), 73–84.

Rand Corporation,
Santi Monica, California

ENUMERATION OF LABELLED GRAPHS

E. N. GILBERT

1. Introduction. The number of connected linear graphs having V vertices labelled $1, \ldots, V$ and λ (unlabelled) lines is found below. Similar formulas are found for graphs in which slings, lines "in parallel," or both are allowed and for directed graphs with or without slings or parallel lines. Some of these graphs are also counted when the lines are labelled and the vertices are unlabelled. Another type of graph which is counted is connected, has no cycles of odd length (even graph), and has L labelled lines and λ unlabelled vertices.

Two graphs with labelled vertices are counted as the same if and only if for all i and j the same number of lines go from the vertex labelled i to the vertex labelled j in both graphs. Consequently two topologically equivalent graphs may be counted as distinct labelled graphs if they are labelled differently. An enumeration of some unlabelled connected graphs has been given by Riddell and Uhlenbeck **(9)** and by Harary **(4)**. Riddell and Uhlenbeck also count connected graphs with labelled vertices in which slings, lines in parallel, and directed lines are ruled out.

2. Graphs with labelled vertices. The kinds of labelled graphs mentioned above are easy to count if one removes the condition that the graph be connected. For example the number of labelled graphs (including disconnected graphs) which have V vertices labelled $1, \ldots, V$ and λ unlabelled lines and no pair of vertices joined by more than one line (no lines in *parallel*) is clearly the binomial coefficient $\binom{\frac{1}{2}V(V-1)}{\lambda}$. For, every such graph is just a collection of λ of the $\frac{1}{2}V(V-1)$ lines which can be drawn between pairs of distinct vertices. More generally let P be any property of connected graphs. Let $T_{V,\lambda}$ be the total number of graphs having V labelled vertices, λ unlabelled lines and such that every connected component of the graph has the property P. Suppose the numbers $T_{V,\lambda}$ are known and consider the number $C_{V,\lambda}$ of connected graphs of V labelled vertices and λ unlabelled lines which have the property P. To get a recurrence equation for $C_{V,\lambda}$ note that in a graph J with $V+1$ vertices and λ lines the vertex labelled $V+1$ belongs to a connected component K having some number v of other vertices and some number μ of lines. The remaining part $J-K$ of J has $V-v$ vertices and $\lambda-\mu$ lines. There are $\binom{V}{v}$ ways in which v of the labels $1, \ldots, V$ can be chosen to be

Received December 5, 1955; in revised form February 6, 1956.

assigned to the graph K, then $C_{v+1,\mu}$ ways of picking K, and $T_{V-v,\lambda-\mu}$ ways of picking $J - K$. Hence we conclude

$$T_{V+1,\lambda} = \sum_{v,\mu} \binom{V}{v} C_{v+1,\mu} T_{V-v,\lambda-\mu}. \tag{1}$$

In (1) we must make the convention that $T_{0,\lambda} = 1$ if $\lambda = 0$ and $T_{0,\lambda} = 0$ otherwise. Introducing the generating functions

$$C_V(y) = \sum_\lambda C_{V,\lambda} y^\lambda$$

and

$$T_V(y) = \sum_\lambda T_{V,\lambda} y^\lambda$$

(1) assumes a simple form

$$T_{V+1}(y) = \sum_v \binom{V}{v} C_{v+1}(y) T_{V-v}(y) \tag{2}$$

which relates $C_{V+1}(y)$ (the term $v = V$) to $C_1(y), \ldots, C_V(y)$ and the known $T_n(y)$.

For computing purposes (2) is quite convenient. However, we will also solve (2) for $C_{V+1}(y)$ explicitly in terms of $T_V(y)$. This solution is obtained in the form of the generating function

$$C(x, y) = \sum_V C_V(y)\, x^V/V!. \tag{3}$$

A compact derivation is achieved using the symbolic method (**2**; **6**). We use a special pair of parentheses $\{\ldots\}_s$ to enclose expressions which are to be interpreted symbolically. The expressions inside such parentheses will be analytic functions depending on x and two letters C and T. The entire expression, including parentheses, stands for the formula which is obtained when the analytic function is expanded into a power series and the term $T^m C^n x^i$ is replaced by $T_m(y)\, C_n(y) x^i$ for all m, n, i. For example (3) becomes

$$C(x, y) = \{\exp Cx\}_s$$

and (2) becomes

$$T_{V+1}(y) = \{C(T + C)^V\}_s. \tag{4}$$

Multiplying both sides of (4) by $x^V/V!$ and summing on V one derives

$$\begin{aligned} \{T \exp Tx\}_s &= \{C \exp (T + C)x\}_s \\ &= \{C \exp Cx\}_s \{\exp Tx\}_s \end{aligned}$$

$$\{C \exp Cx\}_s = \{T \exp Tx\}_s / \{\exp Tx\}_s. \tag{5}$$

Integrating both sides of (5) with respect to x from 0 to x one also has

$$\{\exp Cx\}_s - \{C^0\}_s = \log \{\exp Tx\}_s.$$

If by convention we put $\{C^0\}_s = C_0(y) = 0$, then

$$C(x, y) = \{\exp Cx\}_s = \log \{\exp Tx\}_s. \tag{6}$$

Either of (5) or (6) provides an explicit solution for $C_V(y)$ which, in the usual notation, is given by the following theorem.

THEOREM 1. *Let $T_V(y)$ be the generating function for the number of graphs with V labelled vertices, λ lines, and such that each component has a given property P. Let $C_V(y)$ be the generating function for the number of connected graphs having V labelled vertices, λ lines, and having property P. Then $C_V(y)$ is $(V-1)!$ times the coefficient of x^{V+1} in the power series for the quotient*

$$\sum_{V=0}^{\infty} T_{V+1}(y)\frac{x^V}{V!} \Big/ \sum_{V=0}^{\infty} T_V(y)\frac{x^V}{V!}.$$

$C_V(y)$ is also $V!$ times the coefficient of x^V in the power series for

$$\log \sum_{V=0}^{\infty} T_V(y)\, x^V/V!.$$

In these series $T_0(y) = 1$ by convention.

All the steps leading up to Theorem I can be justified using results in a paper on the symbolic method by Bell (**2**). In all our applications the power series $\{\exp Cx\}$, has a zero radius of convergence; nevertheless $C_V(y)$ is obtained from either form of the generating series by formally expanding the generating series just as though it converged. One can also prove Theorem I using traditional methods but the proof is laborious. For example, expanding the generating function

$$\log\left(1 + \sum_{V=1}^{\infty} T_V(y)\, x^V/V!\right)$$

into a formal power series in x, Theorem I states that

$$C_V(y) = -V! \sum_{b_1,\ldots,b_V} \frac{(\Sigma b_i - 1)!}{b_1!\ldots b_V!}\left(-\frac{T_1(y)}{1!}\right)^{b_1}\ldots\left(-\frac{T_V(y)}{V!}\right)^{b_V} \tag{7}$$

where the sum is taken over all partitions $b_1 + 2b_2 + \ldots + Vb_V = V$. It is then possible to retrace the steps of our symbolic proof backward to show that the expression (7) is indeed the solution of the recurrence equation (2).

It was noted above that there are

$$\binom{\frac{1}{2}V(V-1)}{\lambda}$$

graphs with λ lines and V labelled vertices and no lines in parallel. Hence putting

$$T_V(y) = (1+y)^{\frac{1}{2}V(V-1)} \tag{8}$$

in Theorem I or in (2) we count connected graphs with V labelled vertices and λ lines, none in parallel. Setting $y = 1$ we count these graphs by vertices, allowing any number of lines. For $V = 1, 2, \ldots, 5$ we find 1, 1, 4, 38, and 728 connected graphs. The 38 connected graphs with four labelled vertices fall into the six topologically distinct types shown in Figure 1.

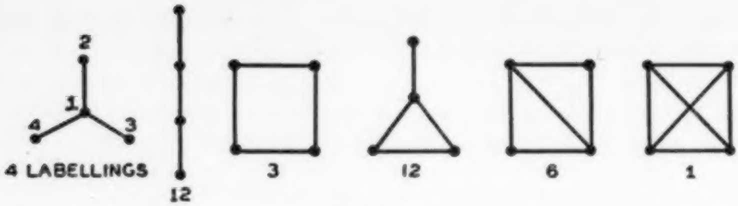

Figure 1

Similarly to count graphs in which any number of lines in parallel are allowed $T_{V,\lambda}$ is the number of combinations of λ lines drawn from $\frac{1}{2}V(V-1)$ different kinds of lines with repetitions allowed, i.e.

$$T_{V,\lambda} = \binom{\frac{1}{2}V(V-1) + \lambda - 1}{\lambda}$$

and

$$(9) \qquad T_V(y) = (1-y)^{-\frac{1}{2}V(V-1)}.$$

A *sling* is a line of a graph which has both its end-points the same. Two slings which share the same vertex will be considered to be in parallel. Slings were ruled out in the above enumerations. If slings are to be allowed the same sort of argument applies. Now there are V additional kinds of lines to choose from so that the terms $\frac{1}{2}V(V-1)$ in (8) and (9) are to be replaced by $\frac{1}{2}V(V+1)$.

Similarly various kinds of directed graphs may be counted. For instance if the graphs are to be composed entirely of lines directed between different vertices but slings are excluded there are $V(V-1)$ kinds of lines and the exponents of (8) and (9) are to be multiplied by 2. Note that this modification of (8) counts two lines joining the same pair of vertices as not in parallel if they have opposite direction. To count directed graphs with slings the exponents in (8) and (9) are changed to V^2, which is the number of kinds of directed lines including the V slings (which can have only one direction). A diadic relation may be interpreted as a directed graph in which there is a line directed from i to j if i has the given relation to j. In a paper enumerating structures of relations, Davis **(3)** counts certain kinds of topologically different (i.e. unlabelled) directed graphs.

Our results are summarized most compactly by the generating functions of the following theorem.

THEOREM II. *The number of connected graphs having V labelled vertices and λ unlabelled lines is $V!$ times the coefficient of $x^V y^\lambda$ in a generating series of the form*

$$\log\left(1 + \sum_{i=1}^{\infty} \frac{(1+\alpha y)^{\alpha\beta(i)} x^i}{i!}\right),$$

where

$$\alpha = \begin{cases} -1 & \textit{if lines in parallel are allowed,} \\ 1 & \textit{otherwise,} \end{cases}$$

and

$$\beta(i) = \begin{cases} \binom{i}{2} & \textit{for (undirected) graphs without slings,} \\ \binom{i+1}{2} & \textit{for (undirected) graphs with slings,} \\ i(i-1) & \textit{for directed graphs without slings,} \\ i^2 & \textit{for directed graphs with slings.} \end{cases}$$

The formula with $\alpha = 1$, $\beta(i) = \binom{i}{2}$ was derived by Riddell and Uhlenbeck **(8)** using quite different techniques.

3. Graphs with labelled lines. Graphs having labelled lines and unlabelled vertices may be counted by exactly the same technique as used in §2. Equation (1) remains true if we interpret $C_{V,\lambda}$ as the number of connected graphs with property P and having V lines labelled $1, 2, \ldots, V$ and λ unlabelled vertices and $T_{V,\lambda}$ as the number of (perhaps disconnected) graphs having V labelled lines, λ unlabelled vertices and such that every component has property P. With this change in viewpoint Theorem I counts connected graphs with labelled lines if the corresponding graphs (not necessarily connected) can be counted.

As an example of such an enumeration, let P be the property that every cycle of the graph must be of even length. König **(7**, Ch. XI, §4**)** calls such a graph an *even graph*. We now use L for the number of labelled lines. Lines in parallel will be permitted but slings are cycles of odd length and so are excluded automatically.

The vertices of even graphs separate into two classes A and B with the property that every line has one end-point in A and one in B. The connections among the L end-points in A and the L end-points in B form two groupings of the L lines into clusters; all lines belonging to the same cluster are connected together at a single vertex. It will be convenient first to count even graphs in which each vertex carries a label A or B to show to which of the two classes it belongs. Let a denote the number of vertices in A and b the number in B. The number of graphs with these values of a and b is just $S(L, a)\ S(L, b)$ where $S(L, k)$ is the number of ways of putting L different objects (the end-points) into k groups (the vertices) so that no group is empty. $S(L, k)$ is a Stirling number of second kind **(5**, p. 179**)** and is given by the generating function

$$\sum_{L,k} S(L,k)\, y^k \frac{x^L}{L!} = \exp[y(\exp x - 1)]. \tag{10}$$

The generating function $T_L(y)$ for the number of graphs with λ vertices is then

$$
\begin{aligned}
T_L(y) &= \sum_{a,\lambda} S(L, a)\, S(L, \lambda - a)\, y^\lambda \\
&= \Big(\sum_k S(L, k)\, y^k\Big)^2.
\end{aligned}
\tag{11}
$$

Theorem I together with (11) solves the problem for even graphs with A, B labels on the vertices. When the A, B labels are removed one finds that most of the even graphs were counted twice (once for each choice of the A, B labelling). However, those graphs for which the A and B partitions were alike were counted just once. Of course there is only one connected even graph of this sort (the graph with two vertices and all L lines in parallel). Hence the generating series $c(x, y)$ for even graphs is one half the sum of the series for even graphs with A, B labels plus the series

$$\sum_{L=1}^{\infty} 1 \cdot y^2 \frac{x^L}{L!} = y^2(\exp x - 1).$$

The final result is

THEOREM III. *The number of even graphs with L lines labelled* $1, 2, \ldots, L$ *and* λ *unlabelled vertices is* $L!$ *times the coefficient of* $x^V y^\lambda$ *in the generating series*

$$\frac{1}{2}\left\{y^2(\exp x - 1) + \log\left(1 + \sum_{k=1}^{\infty} T_k(y)\, x^k/k!\right)\right\}$$

where $T_k(y)$ *is given by* (11).

To count even graphs by lines allowing any number of vertices, set $y = 1$. Then $T_k(1) = G_k^2$ where G_k is the number of ways of grouping k distinct objects into any number of clusters. A table of G_k is given by Bell (**1**). For $L = 1, 2, \ldots, 5$ there are 1, 2, 8, 60, 672 even graphs. The 60 even graphs with 4 labelled lines fall into 10 topologically distinct types shown in Figure 2.

A simple enumeration may be given for connected directed graphs allowing lines in parallel and slings. If we drop the connectedness requirement the number of such graphs which have L labelled lines and λ unlabelled vertices

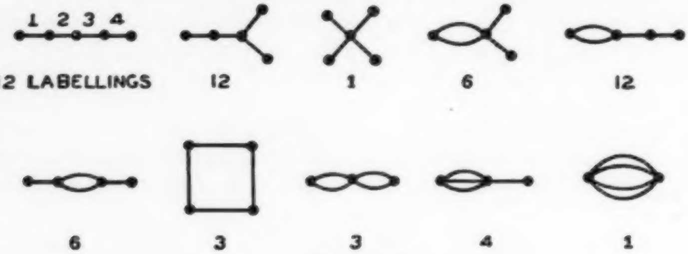

Figure 2

is $S(2L, \lambda)$; for, the vertices represent clusters of end-points taken from the $2L$ distinct end-points of the L directed lines. Theorem I now applies with

$$T_L(y) = \sum_{\lambda} S(2L, \lambda) y^{\lambda}.$$

This enumeration can be modified to rule out slings. The number $S(2L, \lambda)$ must now be replaced by the number of groupings of $2L$ distinct end-points into λ clusters such that no cluster contains both end-points of a line. This number is easily found by the principle of inclusion and exclusion (8)

$$T_{L,\lambda} = \sum_{k} (-1)^k \binom{L}{k} S(2L - k, \lambda).$$

Undirected connected graphs with lines in parallel allowed but with no slings may now be counted by noting that, with one exception, each such graph contributes 2^L directed graphs to the preceding enumeration when the lines are directed in all possible ways. The exception is the graph with two vertices (L lines in parallel), which contributes only 2^{L-1} directed graphs because a simultaneous reversal of all L directions on this graph does not produce a new directed graph.

References

1. E. T. Bell, *Exponential polynomials*, Ann. Math., *35* (1934), 258–277.
2. ———, *Postulational bases for the umbral calculus*, Amer. J. Math., *62* (1940), 717–724.
3. R. L. Davis, *The number of structures of finite relations*, Proc. Amer. Math. Soc., *4* (1953), 486–495.
4. F. Harary, *The number of linear, directed, rooted and connected graphs*, Trans. Amer. Math. Soc., *78* (1955), 445–463.
5. C. Jordan, *Calculus of Finite Differences* (New York, 1950).
6. I. Kaplansky and J. Riordan, *Multiple matching and runs by the symbolic method*, Ann. Math. Stat. (3), *16* (1945), 272–277.
7. D. König, *Theorie der endlichen und unendlichen Graphen* (New York, 1950).
8. E. Netto, *Lehrbuch der Combinatorik* (Leipzig, 1901).
9. R. J. Riddell, Jr. and G. E. Uhlenbeck, *On the theory of the virial development of the equation of state of monoatomic gases*, J. Chem. Phys., *21* (1953), 2056–2064.

Bell Telephone Laboratories,
Murray Hill, N.J.

CERTAIN INFINITE ZERO-SUM TWO-PERSON GAMES

A. L. DULMAGE AND J. E. L. PECK

1. Introduction. The theorem of von Neumann, that every finite, zero-sum two-person game has a value, has been extended in various ways to infinite games. In particular Wald **(6)** has shown that any bounded game in which one player has finitely many pure strategies, has a value. Our interest was aroused by the infinite analogue of the game of "hide and seek" as described by von Neumann **(5)**, which does not appear to fit any of the known cases, unless the matrix is bounded. However, the bounded game is dull since its value is zero. This has led us to give another set of sufficient conditions under which an unbounded infinite game may have a value.

2. Notation and definitions. The game (I, J, K) will consist of two arbitrary sets I and J and a real function K on the product set $I \times J$. If the maximising player chooses $i \in I$ and the minimising player independently chooses $j \in J$, then the former receives the amount $K(i, j)$ from the latter.

Let $\xi = \{x_i; i \in I\}$ denote a vector with dimension the cardinality of I, such that $\sum x_i = 1$ and all $x_i \geqslant 0$, the sum being taken in the sense of Bourbaki **(3**, Ch. III, §4**)**. The vector ξ will be used as a mixed strategy for the maximising player. Similarly $\eta = \{y_j\}$ will be used as a mixed strategy for the minimising player. We write

$$K(\xi, \eta) = \sum_{i,j} K(i, j) x_i y_j$$

when the expression on the right is summable in the sense of Bourbaki **(3)**. For a particular ξ,

$$\inf_{\eta} K(\xi, \eta)$$

will denote the infimum over all those η for which $K(\xi, \eta)$ exists. We write

$$\underline{v}_{IJ} = \sup_{\xi} \inf_{\eta} K(\xi, \eta)$$

and similarly

$$\bar{v}_{IJ} = \inf_{\eta} \sup_{\xi} K(\xi, \eta).$$

If the sets I and J are finite then the fundamental theorem states that $\underline{v}_{IJ} = \bar{v}_{IJ}$. If J is a finite set and the $K(i, j)$ are bounded then it is known **(1)** that $\underline{v}_{IJ} = \bar{v}_{IJ}$. In the general case, if $K(\xi, \eta)$ exists for all ξ and η then

$$\underline{v}_{IJ} \leqslant \bar{v}_{IJ}. \tag{1}$$

Received, October 20, 1955. This work was done while the authors were fellows of the Summer Research Institute of the Canadian Mathematical Congress.

However, as we may see from an example of Bohnenblust, Karlin and Shapley **(2)**, if there is some $K(\xi, \eta)$ which does not exist, we may have $\bar{v}_{IJ} > \underline{v}_{IJ}$.

Games for which (1) holds we call *admissible*, and we say that any game for which $\bar{v}_{IJ} = \underline{v}_{IJ}$ is *determinate*, or has a *value*. Any game for which the $K(i, j)$ are bounded below or bounded above is admissible, because all the $K(\xi, \eta)$ exist. It is a simple matter to construct admissible games and in fact determinate games in which, for some ξ and η, $K(\xi, \eta)$ does not exist. For any admissible game we have from (1) and the definitions that

$$(2) \qquad \underline{v}_{IJ} \leq \bar{v}_{IJ} \leq \bar{v}_{In},$$

for any subset n of J.

If M is the collection of finite subsets of I directed by inclusion, then $\lim_i f(i) = A$ will mean that for all $\epsilon > 0$ there exists an $m \in M$ such that for all $i \notin m$, $|f(i) - A| < \epsilon$. Similarly $\lim_{m \in M} f(m) = A$ will mean that for all $\epsilon > 0$ there exists an $m' \in M$ such that for all $m \supseteq m'$, $|f(m) - A| < \epsilon$. There are obvious modifications in the case where A is not finite. Similarly N will represent the collection of finite subsets of J. We denote the cardinal number of the set s by $|s|$.

3. A sufficient condition for an admissible game to have a value.

THEOREM. *If the game (I, J, K) is admissible and if for each $j \in J$ there is a real number L_j such that*

$$(3) \qquad \inf_i K(i, j) = L_j$$

$$(4) \qquad \lim_i K(i, j) = L_j,$$

then the game has a value

$$v = \lim_{n \in N} \bar{v}_{In}, \qquad v < +\infty,$$

and the maximising player has an optimal strategy.

Proof. For any $n \in N$, the game (I, n, K) is bounded by hypothesis, and so **(1)** has a value $v_{In} = \bar{v}_{In} = \underline{v}_{In}$. But if $n \subseteq n'$, then $v_{In} \geq v_{In'}$ so that we may write

$$(5) \qquad v = \lim_{n \in N} v_{In}$$

and $v < +\infty$. From (2) and (5) we have that

$$(6) \qquad \underline{v}_{IJ} \leq \bar{v}_{IJ} \leq v.$$

If $v = -\infty$, then the game has the value $-\infty$ and every strategy for the maximising player is optimal. Thus we need only consider the case where v is finite.

For each $n \in N$ we may choose a strategy $\xi_n = \{x_{ni}\}$ for the maximising player which is $|n|^{-1}$ optimal for the game (I, n, K) in the sense that for all $j \in n$

$$K(\xi_n, j) = \sum_i K(i,j)x_{ni} \geqslant v_{In} - |n|^{-1}. \tag{7}$$

Since the closed interval [0, 1] is compact, it follows from the Tychonoff theorem that the product π of $|I|$ of these intervals is compact **(4)** in the topology of coordinate-wise convergence. Since for every $n \in N$, $\xi_n = \{x_{ni}\}$ lies in π, the net $\{\xi_n; n \in N\}$ has a convergent **(4)** subnet $\{\xi_n; n \in N'\}$. For each i put

$$x_i' = \lim_{n \in N'} x_{ni}.$$

Then for every $i \in I$

$$0 \leqslant x_i' \leqslant 1, \quad \sum_{i \in I} x_i' \leqslant 1. \tag{8}$$

We now write $\xi' = \{x_i'\}$ even though ξ' may not be a strategy.

Put $L(i,j) = K(i,j) - L_j$, then from (3)

$$L(i,j) \geqslant 0, \tag{9}$$

and from (4),

$$\lim_i L(i,j) = 0. \tag{10}$$

With the obvious interpretation of $L(\xi_n, j)$ and $L(\xi', j)$ we shall prove that, for all $j \in J$,

$$\lim_{n \in N'} L(\xi_n, j) = L(\xi', j). \tag{11}$$

In fact, given $\epsilon > 0$, choose, by (10), a finite subset $m_j \subseteq I$ so that when $i \notin m_j$, $L(i,j) < \frac{1}{3}\epsilon$, and putting

$$B_j = \max_i L(i,j),$$

choose $n' \in N'$ so that whenever n follows n' in N' we have for all $i \in m_j$ that $|x_{ni} - x_i'| < \epsilon(3|m_j|B_j)^{-1}$. Then

$$\begin{aligned}
|L(\xi_n, j) - L(\xi', j)| &= |\sum_i L(i,j)x_{ni} - \sum_i L(i,j)x_i'| \\
&\leqslant \sum_{i \in m_j} L(i,j)|x_{ni} - x_i'| + \sum_{i \notin m_j} L(i,j)x_{ni} + \sum_{i \notin m_j} L(i,j)x_i' \\
&< B_j \sum_{i \in m_j} \epsilon(3|m_j|B_j)^{-1} + \tfrac{1}{3}\epsilon \sum_i x_{ni} + \tfrac{1}{3}\epsilon \sum_i x_i' \\
&\leqslant \tfrac{1}{3}\epsilon + \tfrac{1}{3}\epsilon + \tfrac{1}{3}\epsilon = \epsilon.
\end{aligned}$$

Since for each $j \in n$, we have from (7) that

$$v_{In} - |n|^{-1} \leqslant L(\xi_n, j) + L_j$$

it follows from (5) and (11) that for all $j \in J$

$$v \leqslant L(\xi', j) + L_j. \tag{12}$$

Let $\sum x_i' = \theta$, then from (8) $0 \leqslant \theta \leqslant 1$ and if $\xi^* = \{x_i^*\}$ is any mixed strategy, put $\xi = \xi' + (1 - \theta)\xi^*$, that is for all i, $x_i = x_i' + (1 - \theta)x_i^*$.

From (8) we have that $\sum x_i = 1$ and that all $x_i > 0$ so that ξ is a mixed strategy. Using (12), (9) and the fact that $x_i' < x_i$, we have for all $j \in J$ that

$$v < \sum_i L(i,j)x_i' + L_j < \sum_i L(i,j)x_i + L_j \sum_i x_i = \sum_i K(i,j)x_i = K(\xi,j),$$

and therefore that

$$v < \inf_j K(\xi,j).$$

However, from the definition of $\bar{v}_{IJ}$,

$$\inf_j K(\xi,j) < \bar{v}_{IJ};$$

so, using (6),

$$v < \inf_j K(\xi,j) < \bar{v}_{IJ} < \bar{v}_{IJ} < v.$$

This proves that the game has a value v, which by (5) is

$$\lim_{n \in N} v_{I_n},$$

and that ξ is an optimal strategy for the maximising player.

4. The infinite game of hide and seek. This game is played on a countably infinite matrix (α_{ij}) where $\alpha_{ij} > 0$. The hider chooses a place (i,j) and the seeker chooses either a row i or a column j and if he "finds" the hider, the amount α_{ij} passes from the hider to the seeker. In the finite $n \times n$ case, von Neumann has shown **(5)** that the value of the game is S_n^{-1}, where, if P_n is some permutation of the integers $i = 1, \ldots, n$,

$$S_n = \max_{P_n} \sum_{i=1}^{n} (\alpha_{iP_n})^{-1}.$$

We observe, in both the finite and the infinite case, that corresponding to every pure strategy of the hider, the seeker has only two pure strategies in which the pay-off is positive and in all other cases it is zero. If this game is considered in the normal form (I, J, K), this means that, for each j, every $K(i,j)$ is zero except for two which are positive. The game is clearly admissible, since the $K(i,j)$ are bounded below by zero. The conditions of our theorem are easily satisfied with $L_j = 0$ for all j and the infinite game therefore has the value

$$v = \lim_{n\to\infty} v_n,$$

where v_n is the value of the game (∞, n^2, K). However the game (∞, n^2, K) is clearly equivalent to the game $(2n, n^2, K)$, whose value is S_n^{-1}, because if the hider is restricted to a square, the seeker would not seek outside it. Thus

$$v = \lim_{n\to\infty} v_n = \lim_{n\to\infty} S_n^{-1}.$$

If we let P be any permutation of the set of positive integers onto itself and if

$$S = \sup_P \sum_{i=1}^{\infty} (\alpha_{ii^P})^{-1},$$

then it is easily shown that

$$\lim_{n\to\infty} S_n = S,$$

and therefore that $v = S^{-1}$. In fact, if $S < \infty$, then we can choose P so that

$$\sum_{i=1}^{\infty} (\alpha_{ii^P})^{-1} > S - \tfrac{1}{2}\epsilon,$$

and then choose n so that

$$\sum_{i=1}^{n} (\alpha_{ii^P})^{-1} > S - \epsilon.$$

If

$$m = \max_{1 \leqslant i \leqslant n} i^P,$$

then we have

$$S \geqslant S_m \geqslant \sum_{i=1}^{n} (\alpha_{ii^P})^{-1} > S - \epsilon.$$

The case $S = \infty$ is similar. This justifies our statement in §1 that the value of the bounded game is zero, since in that case $S = \infty$.

There exist unbounded hide and seek games in which the value is not zero, for example if $\alpha_{ij} = 2^{\max(i,j)}$, the value of the game is 1.

There is the same connection between the game of hide and seek and the optimal assignment problem, as in the finite case, but with obvious modifications.

More general games of hide and seek can be considered as played on a t-dimensional array $(\alpha_{i_1,\ldots,i_t})$ where the hider chooses a place $(i_1,\ldots,i_t)$ and the seeker chooses some r subscripts, $r < t$. Our theorem shows that such infinite games have a value.

References

1. D. Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions* (Wiley, 1954).
2. H. F. Bohnenblust, S. Karlin and L. S. Shapley, *Solutions of discrete games*, Ann. Math. Studies, *24* (1950), 51–72.
3. N. Bourbaki, *Topologie générale* (Paris, Hermann).
4. J. L. Kelley, *Convergence in topology*, Duke Math. J., *17* (1950), 277–283.
5. J. von Neumann, *A certain zero-sum two-person game equivalent to the optimal assignment problem*, Ann. Math. Studies, *28* (1953), 5–12.
6. A. Wald, *Generalization of a theorem by v. Neumann concerning zero sum two person games*, Ann. Math., *46* (1945), 281–286.

Royal Military College of Canada *University of New Brunswick*

ON THE BASIS PROBLEM FOR VECTOR VALUED FUNCTION SPACES

H. W. ELLIS

1. Introduction. In a recent paper **(2)** Halperin and the author considered separable Banach spaces L^λ of real valued functions on general measure spaces and proved the existence of 1-regular (§2) Haar or σ-Haar bases when λ was the classical p-norm or any levelling length function **(3)** and, more generally, of K-regular Haar or σ-Haar bases when λ was a continuous length function satisfying certain additional conditions **(2**, Theorem 3.2**)**.

In the present note, separable spaces $L^\lambda(S; X)$, $V^\lambda(S; X)$ of functions valued in a normed vector space X on a general measure space S are considered and the existence of a $3KK'$-regular basis is established when $L^\lambda(V^\lambda)$ has a K-regular Haar or σ-Haar basis and X has a K'-regular basis.

2. Terminology. S will denote an arbitrary space of points P with a countably additive, non-negative measure $\gamma(E)$ defined for a complemented, countably additive family of sets; λ will be an arbitrary length function; $L^\lambda(S)$ will denote the Banach space of real valued functions $f(P)$ on S with $\lambda(f)$ defined and finite; X will denote an arbitrary normed vector space with real scalars; $\|\ \|$ the norm in X; $L^\lambda(S; X)$ the space of Bochner measurable functions **(4)** $f(P)$ valued in X on S with $\lambda(f) = \lambda[f(P)] = \lambda(\|f(P)\|)$ defined and finite. (If X is complete $L^\lambda(S; X)$ is a Banach space **(3)**.)

Upper case letters will be used for arbitrary measurable sets, lower case letters will always denote measurable sets of finite measure; $f_E(P)$ will denote the function equal to $f(P)$ in E and vanishing elsewhere and $\lambda(E)$ will be an abbreviation for $\lambda(1_E)$.

Definition. A basis $\{x_i\}$ in X will be called *K-regular*, if

2.1 $$\left\|\sum_1^n a_i x_i\right\| \leqslant K\|x\|, \quad 1 \leqslant K < \infty, \qquad n = 1, 2, \ldots$$

for every $x = \sum a_i x_i \in X$.

The referee has pointed out that, if X is a Banach space, Banach's boundedness theorem **(1**, p. 80**)** shows that any basis in M is a K-regular basis for some K.

3. Bases in $L^\lambda(S; X)$. Suppose that X has a K-regular basis $\{x_i\}$. If $f(P)$ is valued in X,

3.1 $$f(P) = \sum a_i(P)\, x_i,$$

Received May 26, 1955.

the $a_i(P)$ being determined uniquely for each i and all P. We shall show that if $f(P) \in L^\lambda(S; X)$ then the real valued functions $a_i(P)$ $(i = 1, 2, \ldots)$ belong to $L^\lambda(S)$. This involves showing that each $a_i(P)$ is measurable with $\lambda[a_i(P)] < \infty$. We note that the uniqueness of the representation $x = \sum a_i x_i \in X$ implies that if $f(P)$ is constant in E so is each $a_i(P)$. Since

$$3.2 \quad |a_i(P)| \, ||x_i|| < \left|\left| \sum_{j=1}^{i} a_j(P) x_j \right|\right| + \left|\left| \sum_{j=1}^{i-1} a_j(P) x_j \right|\right| < 2K||f(P)||,$$

and $f(P)$ is the almost uniform limit of countably valued functions, each $a_i(P)$ is the almost uniform limit of measurable (countably valued) functions and is therefore measurable. Thus $\lambda(a_i)$ is defined for each i. Using (3.2) and properties (L 2) and (L 4) of length functions **(3)**,

$$\lambda[a_i(P)] = \lambda(|a_i(P)|) < 2K\lambda(||f(P)||)/||x_i|| < \infty.$$

Lemma 3.1. *If $f(P) \in L^\lambda(S; X)$, where λ is a continuous length function and $L^\lambda(S)$ is separable, if X has a K'-regular basis $\{x_i\}$, and if* (3.1) *holds, then*

$$3.3 \qquad \lim_{n\to\infty} \lambda\left[f(P) - \sum_1^n a_i(P) x_i\right] = 0.$$

Proof. First suppose that $\gamma(S) < \infty$. Given $\epsilon > 0$, let e denote the set of points P for which

$$\left|\left| f(P) - \sum_1^n a_i(P) x_i \right|\right| < \epsilon$$

for all $n > N$. Then $\gamma(S - e) \to 0$ as $N \to \infty$,

$$\lambda\left[f(P) - \sum_1^n a_i(P) x_i\right] < \epsilon\gamma(S) + (1 + K')\,\lambda(f_{S-e}),$$

and $\lambda(f_{S-e}) \to 0$ as $N \to \infty$ by **(2**, Lemma 3.2**)**.

If S is arbitrary there exists e' by **(2**, Lemma 3.2 (iii)**)** with $\lambda(f - f_{e'})$ arbitrarily small,

$$\lambda\left[f(P) - \sum_1^n a_i(P) x_i\right] < (1 + K')\,\lambda(f - f_{e'}) + \lambda\left[\left(f - \sum_1^n a_i(P) x_i\right)_{e'}\right]$$

and the right side can be made arbitrarily small by choice of e' and n.

Lemma 3.2. *Let $f(P) \in L^\lambda(S; X)$, let $L^\lambda(S)$ have a Haar or σ-Haar basis and X have a K'-regular basis $\{x_i\}$ and suppose that* (3.1) *holds. Then* (3.3) *holds.*

Proof. The assumption that $L^\lambda(S)$ has a σ-Haar basis implies that $L^\lambda(S)$ is separable and that $S = E + \bigcup e_n$ where $\lambda(f_E) = 0$ for every $f(P) \in L^\lambda(S)$ and where the σ-Haar basis functions correspond to a σ-Haar system of sets $H_\sigma(\bigcup e_n)$ which forms a countable basis **(2)** in S. Then for arbitrary $f(P) \in L^\lambda(S)$, where

$$f_N(P) = \begin{cases} f(P) \text{ in } \bigcup_1^N e_n, \\ 0 \text{ elsewhere,} \end{cases}$$

$$\lambda(f) = \lambda(f_{S-E}) = \sup_{N\to\infty} \lambda(f_N),$$

by (L 5) for length functions so that λ is continuous and Lemma 3.1 applies. A similar argument applies if $L^\lambda(S)$ has a Haar basis.

COROLLARY. *Under the hypotheses of Lemma* 3.2, $\lambda[a_n(P)] \to 0$ *as* $n \to \infty$.

If $\{\phi_j(P)\}$ is a basis in $L^\lambda(S)$, $a_i(P) = \sum_j a_{ij}\phi_j(P)$, $(i = 1, 2, \ldots)$, the coefficients a_{ij} being uniquely determined. Thus with each $f(P) \in L^\lambda(S; X)$ can be associated a unique double series

3.4 $$f(P) \backsim \textstyle\sum_i [\sum_j a_{ij}\phi_j(P)]\, x_i.$$

We shall show that the $\{x_i\phi_j(P)\}$, ordered suitably into a single sequence, form a basis in $L^\lambda(S; X)$.

In the proof of the next lemma the Bochner integral would be used if X were a Banach space. To extend the proof to an arbitrary normed vector space X, we generalize the Bochner integral for certain functions by defining, where $x_i \in X$, $g_i(P) \in L(S)$ $(i = 1, 2, \ldots, n)$,

$$(X)\int_S \Big[\sum_1^n x_i g_i(P)\Big] d\gamma(P) = \sum_1^n x_i \int_S g_i(P) d\gamma(P).$$

We shall use the fact that

3.5 $$\left|\left|(X)\int_S \Big[\sum_1^n x_i g_i(P)\Big] d\gamma(P)\right|\right| < \int_S \left|\left|\sum_1^n x_i g_i(P)\right|\right| d\gamma(P).$$

This is easily shown if the g_i are finitely valued, constant in the same sets, and the general result is then obtained by standard arguments.

LEMMA 3.3. *If* $\{\phi_i(P)\}$ *is a K-regular Haar or* σ*-Haar basis in* $L^\lambda(S)$, *if* $\{x_i\}$ *is a* K'*-regular basis in* X, *then for all* m, n,

3.6 $$\lambda\Big[\sum_{i=1}^m \sum_{j=1}^n a_{ij}x_i\phi_j(P)\Big] < KK'\lambda(f).$$

Proof. By (**2**, Lemma 2.1, Corollary 1)

$$\sum_{j=1}^n a_{ij}\phi_j(P) = \sum_{r=1}^N \left(\Big[\gamma(e_r)^{-1}\int_{e_r} a_i(P)\, d\gamma(P)\Big] \text{in } e_r\right)$$

for some sequence of Haar or σ-Haar sets e_r depending only on n where

$$\bigcup_1^N e_r = S - E$$

(E defined as in Lemma 3.2) if $L^\lambda(S)$ has a Haar basis and where

$$\sum_1^n a_{ij}\phi_j(P)$$

vanishes outside $\bigcup e_r$ if $L^\lambda(S)$ has a σ-Haar basis. Let $\|f(P)\| = \sum b_j\phi_j(P)$. Then

$$\begin{aligned}
\lambda\left\{\sum_{i=1}^m \sum_{j=1}^n a_{ij}x_i\phi_j(P)\right\} &= \lambda\left\{\sum_{i=1}^m x_i \sum_{r=1}^N \left(\left[\gamma(e_r)^{-1}\int_{e_r} a_i(P)\,d\gamma(P)\right]\text{in } e_r\right)\right\} \\
&= \left\{\sum_{r=1}^N \left(\left[\gamma(e_r)^{-1}(X)\int_{e_r}\left(\sum_{i=1}^m a_i(P)x_i\right)d\gamma(P)\right]\text{ in } e_r\right)\right\} \\
&< \lambda\left\{\sum_{r=1}^N \left(\left[\gamma(e_r)^{-1}\int_{e_r}\left|\left|\sum_{i=1}^m a_i(P)x_i\right|\right|d\gamma(P)\right]\text{ in } e_r\right)\right\} \\
&< K'\lambda\left\{\sum_{r=1}^N \left(\left[\gamma(e_r)^{-1}\int_{e_r}\|f(P)\|\,d\gamma(P)\right]\text{in } e_r\right)\right\} \\
&= K'\lambda\left\{\sum_{j=1}^n b_j\phi_j(P)\right\} \\
&< KK'\lambda\{f(P)\}.
\end{aligned}$$

The sequence

$$\sum_1^n \sum_1^n a_{ij}x_i\phi_j(P)$$

is KK'-regular in $L^\lambda(S;X)$ and suggests that the $x_i\phi_j(P)$ be ordered so as to give partial sums differing as little as possible from square or rectangular sums. To this end we order then as follows: $x_1\phi_1, x_1\phi_2, x_2\phi_1, x_2\phi_2, \ldots, x_{n-1}\phi_{n-1}, x_n\phi_1, x_n\phi_2, \ldots, x_n\phi_{n-1}, x_1\phi_n, x_2\phi_n, \ldots, x_n\phi_n, \ldots$. Given $f(P) \in L^\lambda(S;X)$, consider 3.4 and let $S_N = S_N(f)$ denote the sum of the first N terms $a_{ij}x_i\phi_j(P)$ with the above ordering. Let

$$S_{M,N} = \sum_{i=1}^M \sum_{j=1}^N a_{ij}x_i\phi_j(P).$$

Then

$$\begin{aligned}
3.7\quad S_N &= \sum_{i=1}^{n-1}\sum_{j=1}^{n-1} a_{ij}x_i\phi_j(P) + \sum_{j=1}^{N-(n-1)^2} a_{nj}x_n\phi_j(P), &(n-1)^2 < N < n(n-1) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n-1} a_{ij}x_i\phi_j(P) + \sum_{i=1}^{N-n(n-1)} a_{in}x_i\phi_n(P), &n(n-1) < N < n^2.
\end{aligned}$$

$$\begin{aligned}
3.8\quad S_N &= S_{n-1,n-1} + S_{n,N-(n-1)^2} - S_{n-1,N-(n-1)^2}, &(n-)^2 < N < n(n-1); \\
&= S_{n,n-1} + S_{N-n(n-1),n} - S_{N-n(n-1),n-1}, &n(n-1) < N < n^2.
\end{aligned}$$

From 3.7, 3.8 and Lemma 3.3 we obtain for every $f \in L^\lambda(S;X)$ and all N

$$3.9\qquad \lambda[S_N(f)] < 3KK'\lambda(f).$$

THEOREM 3.1. *If $L^\lambda(S)$ has a K-regular Haar or σ-Haar basis $\{\phi_j(P)\}$ and if X has a K'-regular basis $\{x_i\}$ then the $x_i\phi_j(P)$ with the ordering of the preceding paragraph form a 3KK'-regular basis in $L^\lambda(S; X)$.*

Proof. We shall give the proof where $L^\lambda(S)$ has a σ-Haar basis, the proof where there is a Haar basis being similar and simpler. We can suppose that the σ-Haar basis functions correspond to a σ-Haar system of sets $H_\sigma(S)$ dense in S (cf. Lemma 3.2). As in Lemma 3.2 each $f \in L^\lambda(S; X)$ is the strong limit of functions f_e (vanishing outside sets of finite measure). Each f_e is the almost uniform limit of finitely valued functions and an easy computation using (2, Lemma 3.2) shows that these functions converge strongly to f_e in $L^\lambda(S; X)$. Finally each set of constancy of an arbitrary finitely valued function g can be approximated arbitrarily closely by finite collections of sets of $H_\sigma(S)$ with the corresponding functions converging strongly to g. We conclude that finitely valued functions with sets of constancy in $H_\sigma(S)$ are dense in $L^\lambda(S; X)$. Since

$$\begin{aligned}\lambda[f(P) - S_N(f)] &< \lambda[f(P) - h(P)] + \lambda[h(P) - S_N(h)] + \lambda[S_N(f) - S_N(h)]\\ &< (1 + 3KK')\,\lambda(f - h) + \lambda[h - S_N(h)],\end{aligned}$$

for any f, h in $L^\lambda(S; X)$ it will be sufficient to prove that $\lambda[f - S_N(f)] \to 0$ as $n \to \infty$ where $f(P)$ is a finitely valued function with sets of constancy in $H_\sigma(S)$. Then, if $f(P) = \sum a_i(P)x_i$, each $a_i(P)$ is finitely valued with the same sets of constancy as $f(P)$ and by (2, Lemma 2.1, Corollary 3) there exists n_0 with

$$\sum_{j=1}^{n} a_{ij}\phi_j(P) = a_i(P), \qquad n > n_0,\ i = 1, 2, \dots .$$

Then

$$\begin{aligned}\lambda[f - S_N(f)] &< \lambda\left[f - \sum_1^{n-1} a_i(P)x_i\right] + \lambda\left(\sum_{i=1}^{n-1} x_i\left[a_i(P) - \sum_{j=1}^{n-1} a_{ij}\phi_j(P)\right]\right)\\ &\quad + \lambda\left[x_n \sum_{j=1}^{N-(n-1)^2} a_{nj}\phi_j(P)\right] \qquad (n-1)^2 < N < n(n-1),\\ &= \lambda_1 + \lambda_2 + \lambda_3\\ &< \lambda\left[f - \sum_1^{n} a_i(P)x_i\right] + \lambda\left(\sum_{i=1}^{N-n(n-1)} x_i\left[a_i(P) - \sum_{j=1}^{n} a_{ij}\phi_j(P)\right]\right)\\ &\quad + \lambda\left(\sum_{i=N-n(n-1)+1}^{N} x_i\left[a_i(P) - \sum_{j=1}^{n-1} a_{ij}\phi_j(P)\right]\right) \quad n(n-1) < N < n^2,\\ &= \lambda_1' + \lambda_2' + \lambda_3';\end{aligned}$$

$\lambda_1 \to 0$, $\lambda_1' \to 0$ as $n \to \infty$ by Lemma 3.2, λ_2, λ_2' and λ_3' vanish if $n - 1 > n_0$ and $\lambda_3 < K\|x_n\|\,\lambda[a_n(P)] \to 0$ as $n \to \infty$ by Lemma 3.2, Corollary.

Write $\sum'$ for sums of terms $a_{ij}x_i\phi_j(P)$ ordered as in Theorem 3.1. We have established the existence of a convergent series $\sum' a_{ij}x_i\phi_j(P)$ with sum $f(P)$

for every $f(P) \in L^\lambda(S; X)$. The $x_i\phi_j(P)$ with the specified ordering will be a basis in $L^\lambda(S; X)$ if there is only one such series for each $f(P)$ and that this is true is a consequence of the uniqueness of the respective series for elements of X and $L^\lambda(S)$ in terms of the bases $\{x_i\}$, $\{\phi_j(P)\}$.

The referee has observed that a K-regular basis in X is a basis in the completion of X, so that there would be no loss of generality in assuming throughout that X is a Banach space. This would permit the use of the Bochner integral in Lemma 3.3.

4. Bases in $V^\lambda(S; X)$. In §3 the assumption that λ is a length function implies that $L^\lambda(S)$ is a Banach space. The above arguments remain valid for more general function spaces. Consider a general normed vector space of measurable functions $f(P)$ with norm $\lambda(f) = \lambda(|f(P)|)$. The definition of a norm implies properties (L 1), (L 3) and (L 4) of length functions for λ. Property (L 2) has played a fundamental role in the proofs in §3. However if the normed vecor space is not required to be complete the results in §3 can be obtained with (L 5) replaced by weaker assumptions.

Suppose that for every measurable function u with $0 < u(P) < \infty$ for almost all P, $\lambda(u)$ is defined with $0 < \lambda(u) < \infty$ and satisfies (L 1)-(L 4) for length functions,

(L 5′) If e is fixed, $e' \subset e, \lambda(f) < \infty$, then $\lambda(f_e) - \lambda(f_{e'}) \to 0$ as $\gamma(e - e') \to 0$,

and

(L 6) $\lambda(u) = \sup_e \lambda(u_e)$ (i.e. λ is continuous).

Let $V^\lambda(S)$ denote the space of real valued functions $f(P)$ with $|f(P)|$ measurable and $\lambda(f) = \lambda(|f|) < \infty$ and let $V^\lambda(S; X)$ be the analogue of $L^\lambda(S; X)$. $V^\lambda(S; X)$ is a normed vector space. If $V^\lambda(S)$ is separable and $\lambda(f) < \infty$, the argument of (**2**, Lemma 3.2) gives:

(i) $\lambda(f_e) \to 0$ as $\gamma(e) \to 0$, and

(ii) there exists e with $\lambda(f - f_e)$ arbitrarily small. With L^λ replaced by V^λ Lemmas 3.1-3.3 and Theorem 3.1 then hold.

References

1. Stefan Banach, *Théorie des opérations linéaires* (Warsaw, 1932).
2. H. W. Ellis and Israel Halperin, *Haar functions and the basis problem for Banach spaces*, J. London Math. Soc., *31* (1956), 28–39.
3. ———, *Function spaces determined by levelling length functions*, Can. J. Math., *5* (1953), 576–592.
4. Einar Hille, *Functional analysis and semi-groups*, Amer. Math. Soc. Coll. Publications, Vol. XXXI (New York, 1948).

Queen's University

ON EXPLICIT BOUNDS IN LANDAU'S THEOREM

J. A. JENKINS

1. The theorem of Landau in question may be stated in the form that if the function $F(Z)$ is regular for $|Z| < 1$ and does not take the values 0 and 1, while

$$F(Z) = a_0 + a_1 Z + \dots$$

is its Taylor expansion about $Z = 0$, then $|a_1|$ has a bound depending only on a_0. In fact $|a_1|$ has a bound depending only on $|a_0|$ and Hayman **(1)** gave the explicit bound

$$|a_1| < 2\,|a_0|\{|\log|a_0|| + 5\pi\}.$$

In a recent paper **(2)** I gave a simple method for obtaining explicit bounds in Schottky's Theorem and applied it also to improving the above bound to

$$|a_1| < 2\,|a_0|\{|\log|a_0|| + 7.77\}.$$

Since writing that paper I have observed that by relatively small modifications of the argument that bound can still be substantially improved.

2. It is well known that, for a given a_0, the maximum value of $|a_1|$ is attained for the function $F_0(Z)$ mapping $|Z| < 1$ onto the universal covering surface of the finite W-plane punctured at 0 and 1 and taking the value a_0 at $Z = 0$. Now $|Z| < 1$ is mapped conformally onto $\Re z > 0$ in such a way that if the mapping function is $Z = Z(z)$ and we set $F_0(Z(z)) = f(z)$, then for a suitable branch of $\log f(z)$ the mapping

$$w = \log f(z) \mp \pi i$$

(where $-$ or $+$ is chosen according as $\Im a_0 > 0$ or $\Im a_0 < 0$) carries the domain determined by the inequalities

$$-\pi < \Im z < \pi, \quad \Re z > 0, \quad |z - \tfrac{1}{2}\pi i| > \tfrac{1}{2}\pi, \quad |z + \tfrac{1}{2}\pi i| > \tfrac{1}{2}\pi$$

onto the strip

$$-\pi < \Im w < \pi$$

so that the boundary points $\pm\pi i$ correspond to themselves. Further, the boundary points of these domains at infinity in whose neighborhoods $\Re z$, $\Re w$ become large and positive correspond and the boundary point $z = 0$ corresponds to the point at infinity in whose neighborhood $\Re w$ becomes large and negative. We denote the point in the z-plane corresponding to $Z = 0$ by b. Moreover we set $\zeta = e^{-z}$, $\omega = e^{-w}$ and denote the corresponding mapping

Received August 10, 1955.

between these planes by $\zeta = \phi(\omega)$ or $\omega = \psi(\zeta)$. The function $\phi(\omega)$ is regular and univalent for $|\omega| < 1$ with $\phi'(0) = 1/16$.

Next we observe that, as was proved in (**2**, p. 80), in obtaining a bound of the form

$$|a_1| < 2\,|a_0|\{|\log|a_0|| + K\},$$

it is enough to confine ourselves to the situation $|a_0| > 1$, $|a_0 - 1| > 1$. Then we use distinct arguments according as $|a_0|$ is near 1 or bounded from 1.For $|a_0|$ near 1 we use the fact that under the mapping from the z-plane to the w-plane the half-plane $\Re z > \frac{1}{2}\pi$ is mapped into the w-plane slit along the half-infinite segments $\Im w = (2n + 1)\pi$, $\Re w < 0$, n running through all integers. Comparing the inner radii of these domains with respect to b and its image with the derivative of the mapping function, namely $a_1/2a_0\Re b$, we get the bound (**2**, p. 81)

$$|a_1| < 2(|a_0|\,|a_0 - 1|)^{\frac{1}{2}} \log|2a_0 - 1 + 2\{a_0(a_0 - 1)\}^{\frac{1}{2}}|\ \Re b/(\Re b - \tfrac{1}{2}\pi).$$

Since the conditions $|a_0| > 1$, $|a_0 - 1| > 1$ imply $\Re b > \frac{1}{6}3^{\frac{1}{2}}\pi$ we have for $|a_0| = t$, $t > 1$

$$\begin{aligned}|a_1| &< (3 + 3^{\frac{1}{2}})\,|a_0|(1 + t^{-1})^{\frac{1}{2}} \log[2t + 1 + 2(t^2 + t)^{\frac{1}{2}}]\\ &< 2|a_0|\{\log|a_0| + \Lambda(t)\}\end{aligned}$$

where

$$\Lambda(t) = \tfrac{1}{2}(3 + 3^{\frac{1}{2}})(1 + t^{-1})^{\frac{1}{2}} \log[2t + 1 + 2(t^2 + t)^{\frac{1}{2}}] - \log t.$$

Unlike the function $L(t)$ used previously the function $\Lambda(t)$ is not monotone increasing. However direct calculation shows that on the range $t > 1$ it first decreases to a minimum and from then on increases. Thus, on an interval $1 < t < t_0$, $\Lambda(t)$ does not exceed the larger of $\Lambda(1)$ and $\Lambda(t_0)$. It proves advantageous to take the interval $1 < t < 1.84$. We readily find

$$\Lambda(1) < 5.90,\quad \Lambda(1.84) < 5.94.$$

Thus for $1 < t < 1.84$, $\Lambda(t) < 5.94$ and for $1 < |a_0| < 1.84$ we have

$$|a_1| < 2|a_0|\{\log|a_0| + 5.94\}.$$

Now we apply to the function $\phi(\omega)$ instead of the bound previously used (**2**, p. 81) the result due to Robinson (**3**, p. 444)

$$\left|\frac{d\zeta}{d\omega}\right| > 16|\zeta|^2\,\frac{1 - |\omega|^2}{|\omega|^2}.$$

Using the fact that

$$\phi'(-a_0^{-1}) = -\frac{2a_0^2\Re b}{a_1 e^b}$$

we get

$$|a_1| < \tfrac{1}{8}\,e^{\Re b}\,\Re b\,\frac{|a_0|^2}{|a_0|^2 - 1}.$$

Moreover (**2**, p. 79)

$$e^{\pi R_0} < 16|a_0| + 8,$$

so

$$|a_1| < (2|a_0| + 1) \log(16|a_0| + 8) \frac{|a_0|^2}{|a_0|^2 - 1}.$$

Then for $|a_0| = t$, $t > 1$, we have

$$|a_1| < 2|a_0|\{\log|a_0| + M(t)\},$$

where

$$M(t) = (t + \tfrac{1}{2})t\,(t^2 - 1)^{-1} \log(16t + 8) - \log t.$$

Direct calculation shows that $M(t)$ is decreasing for $t > 1$. Now $M(1.84) < 5.93$. Thus for $|a_0| > 1.84$ we have

$$|a_1| < 2|a_0|\{\log|a_0| + 5.93\}.$$

Combining this with our previous estimate we have

THEOREM 1. *If $F(Z)$ is regular for $|Z| < 1$, does not take the values* 0 *and* 1 *and has Taylor expansion about $Z = 0$*

$$F(Z) = a_0 + a_1 Z + \dots,$$

then

$$|a_1| < 2|a_0|\{|\log|a_0|| + 5.94\}.$$

As Hayman has remarked in his review of (**2**) (Mathematical Reviews, **16** (1955), 579) the value 5.94 cannot be replaced by 4.37.

REFERENCES

1. W. K. Hayman, *Some remarks on Schottky's Theorem*, Proc. Cambridge Phil. Soc., *43* (1947), 442–454.
2. J. A. Jenkins, *On explicit bounds in Schottky's Theorem*, Can. J. Math., *7* (1955), 76–82.
3. R. M. Robinson, *Bounded univalent functions*, Trans. Amer. Math. Soc., *52* (1942), 426–449.

University of Notre Dame

THE CAUCHY PROBLEM FOR LINEAR PARTIAL DIFFERENTIAL EQUATIONS WITH RESTRICTED BOUNDARY CONDITIONS

E. P. MILES, JR. AND ERNEST WILLIAMS

We shall discuss solutions of linear partial differential equations of the form

$$\Phi(D, x_1, x_2, \ldots x_n)\, u + \Psi(D, t)\, u = 0, \tag{1}$$

where Ψ is an ordinary differential operator of order s with respect to t. Our first theorem gives a solution of (1) for the Cauchy data;

$$\begin{aligned} u(x_1, x_2, \ldots x_n, 0) &= P(x_1, x_2, \ldots x_n), \\ \frac{\partial^j u}{\partial t^j}(x_1, x_2, \ldots x_n, 0) &= 0, \qquad j = 1, 2, \ldots, s-1, \end{aligned} \tag{2}$$

whenever the function P is annihilated by a finite iteration of the operator Φ. This situation occurs if P is a polynomial and Φ any differential operator with constant coefficients and no constant term or if P is polyharmonic and Φ the Laplacian operator. The solution hinges upon the integration of a finite system of ordinary differential equations.

THEOREM 1. *Suppose for some integer k we have*

$$\Phi^k(P) \neq 0, \quad \Phi^{k+1}(P) = 0; \tag{3}$$

urther suppose that $u_0, u_1, \ldots u_k$ are a set of solutions of the system of ordinary differential equations

$$\begin{aligned} \Psi(u_j) + u_{j-1} &= 0, \qquad j = 1, 2, \ldots, k, \\ \Psi(u_0) &= 0, \end{aligned} \tag{4}$$

with initial conditions

$$\begin{aligned} u_0(0) = 1, \quad u_j(0) &= 0, \qquad j > 1, \\ \frac{d^m u_j}{dt^m}(0) &= 0, \qquad m = 1, 2, \ldots, s-1, \text{ all } j; \end{aligned} \tag{5}$$

then a solution of (1) *satisfying* (2) *is*

$$u(x_1, x_2, \ldots x_n, t) = \sum_{j=0}^{k} \Phi^j(P) \cdot u_j. \tag{6}$$

Received July 25, 1955.

Proof.

$$\begin{aligned}[\Phi + \Psi]u &= [\Phi + \Psi]\sum_{j=0}^{k} \Phi^j(P)\cdot u_j \\ &= \sum_{j=0}^{k} [\Phi + \Psi]\, \Phi^j(P)\cdot u_j \\ &= P\cdot[\Psi\,(u_0)] + \sum_{j=1}^{k} \Phi^j(P)[\Psi(u_j) + u_{j-1}] + \Phi^{k+1}(P)\cdot u_k \\ &= 0,\end{aligned}$$

by (3), (4) and the linearity of the operators Φ and Ψ. The conditions (5) on the u_j ensure that (6) satisfies conditions (2) and the proof of the theorem is complete.

As an application of this theorem we construct a solution of

$$a\frac{\partial u}{\partial x} + b\frac{\partial u}{\partial y} + c\frac{\partial u}{\partial t} = 0,\ u(x, y, 0) = x^2y. \tag{7}$$

We note that $\Phi(P) = 2axy + bx^2$, $\Phi^2(P) = 4abx + 2a^2y$, $\Phi^3(P) = 6a^2b$ and $\Phi^4(P) = 0$.

$$u_0 = 1,\ u_1 = \frac{-t}{c},\ u_2 = \frac{t^2}{2c^2},\ u_3 = \frac{-t^3}{6c^3}$$

and

$$\begin{aligned} u(x, y, t) &= x^2y + (2axy + bx^2)\left(\frac{-t}{c}\right) + (4abx + 2a^2y)\left(\frac{t^2}{2c^2}\right) + 6a^2b\left(\frac{-t^3}{6c^3}\right) \\ &= \frac{(cx - at)^2(cy - bt)}{c^3}. \end{aligned} \tag{8}$$

The last form of (8) may also be obtained from the general solution of (7), $F(cx - at, cy - bt)$, by requiring that it reduce to x^2y when $t = 0$.

An alternate set of Cauchy data frequently encountered for (1) when Ψ is a second order operator is

$$u(x_1, x_2, \ldots x_n, 0) = 0, \qquad u_t(x_1, x_2, \ldots x_n, 0) = Q(x_1, x_2, \ldots x_n). \tag{9}$$

For this case an analogous theorem holds:

THEOREM 2. *Suppose* (3) *holds and* $V_0, V_1, \ldots V_k$ *are the solutions of*

$$\Psi(V_0) = 0,\quad \Psi(V_j) + V_{j-1} = 0, \qquad j = 1, 2, \ldots k, \tag{10}$$

with initial conditions

$$\begin{aligned} V_j(0) &= 0, && j = 0, 1, \ldots k, \\ V'_0(0) = 1,\quad V'_j(0) &= 0, && j = 1, 2, \ldots k; \end{aligned} \tag{11}$$

then the Cauchy problem for (1) *with boundary values* (9) *has a solution*

$$u = \sum_{j=0}^{k} \Phi^j(Q)\, V_j. \tag{12}$$

The authors have recently obtained basic sets of homogeneous polynomial solutions **(1)** for the Laplace and wave equations in k variables. Although Theorems 1 and 2 were not discovered until after these basic sets were developed, they provide a natural way for deriving them. Let us consider the wave equation in three space variables and one time variable,

$$\nabla^2 u - \frac{\partial^2 u}{\partial t^2} = 0. \tag{13}$$

A general homogeneous polynomial W, of degree n in (x, y, z, t), has $\binom{n+3}{3}$ arbitrary coefficients. Requiring W to be a solution of (13) gives $\binom{n+1}{3}$ independent[1] conditions on these coefficients showing that

$$\binom{n+3}{3} - \binom{n+1}{3} = (n+1)^2$$

of them are independent. We may construct our basic set of $(n+1)^2$ homogeneous polynomial solutions of degree n for (13) by first applying Theorem 1 to each of the $\binom{n+2}{2}$ monomials $P(x, y, z)$ of type $x^a y^b z^c$, $a + b + c = n$ and then applying Theorem 2 to each of the $\binom{n+1}{2}$ monomials $Q(x, y, z)$ of type $x^\alpha y^\beta z^\gamma$, $\alpha + \beta + \gamma = n - 1$. The resulting homogeneous polynomial solutions of (13) are

$$\binom{n+2}{2} + \binom{n+1}{2} = (n+1)^2$$

in number. The polynomials generated by Theorem 1 contain only one term of degree less than 2 in t, the generator term $x^a y^b z^c$; likewise those generated by Theorem 2 contain only one term $x^\alpha y^\beta z^\gamma\, t$ of lowest degree in t. Thus our set of solutions is independent and, since it is correctly numbered, is a basic set of solutions.

The basic set constructed in the above manner may be represented as follows: for each set of non-negative integers a, b, c, d, $a + b + c + d = n$, $d < 1$

$$W_{a,b,c,d}(x, y, z, t) = \sum_{j=0}^{[\frac{1}{2}n]} \nabla^{2j}(x^a y^b z^c) \cdot \frac{t^{2j+d}}{(2j+d)!} \tag{14}$$

$$= \sum \frac{a!\, b!\, c! [\tfrac{1}{2}D]\, x^A y^B z^C t^D}{\left(\frac{a-A}{2}\right)!\left(\frac{b-B}{2}\right)!\left(\frac{c-C}{2}\right)!\, A!\, B!\, C!\, D!}$$

[1]Independence of these conditions may be readily established by a generalization of Whittaker's footnote on the corresponding harmonic polynomials (Whittaker & Watson, *Modern Analysis*, 4th ed., p. 389) or, as suggested by the referee, by using an argument like that in Courant-Hilbert, vol. I, English ed., at the bottom of p. 512.

where the final summation extends over all non-negative A, B, C and D such that $A+B+C+D=n$, $A\equiv a$, $B\equiv b$, $C\equiv c$, $D\equiv d \bmod 2$, and $A<a$, $B<b$, $C<c$. The final form of (14) corresponds, except for a constant factor, to the basic set for the wave equation developed in **(1)**, but the intermediate form given here for the first time is much simpler to describe and use.

Instead of trying to solve Cauchy's problem for the wave equation in terms of the basic set of polynomials (14) when our initial conditions are suitable we may use:

THEOREM 3. *The solution of* (13) *with initial conditions*

$$u(x, y, z, 0) = P(x, y, z), \quad u_t(x, y, z, 0,) = Q(x, y, z),$$

where $P(x, y, z)$ *and* $Q(x, y, z)$ *are polyharmonic functions of order* p *and* q *respectively, is given by the sum:*

$$(15) \qquad U(x, y, z, t) = \sum_{j=0}^{p-1} \nabla^{2j}(P)\frac{t^{2j}}{(2j)!} + \sum_{k=0}^{q-1} \nabla^{2k}(Q)\frac{t^{2k+1}}{(2k+1)!}.$$

This theorem is established by applying Theorem 1 to (13) with boundary conditions, $u(x, y, z, 0) = P$, $u_t(x, y, z, 0) = 0$ and Theorem 2 to (13) with boundary conditions, $u(x, y, z, 0) = 0$, $u_t(x, y, z, 0) = Q$, and adding the resulting solutions.

As an illustration of Theorem 3 we display a solution of the wave equation (13), with initial conditions

$$(16) \quad u(x, y, z, 0) = xe^x \cos y \cdot z^4, \quad u_t(x, y, z, 0) = 0,$$

$$(17) \quad u(x, y, z, t) = e^x \cos y[xz^4 + (z^4 + 6xz^2)\, t^2 + (2z^2 + x)t^4 + \tfrac{1}{5}t^6].$$

In another recent paper **(2)** the authors gave basic sets of polynomial solutions for the Euler-Poisson-Darboux equation

$$(18) \qquad \nabla^2 u - [u_{tt} + kt^{-1}u_t] = 0$$

and for the closely associated Beltrami equation. For the E.P.D. equation with $k > 0$ a direct application of Theorem 1 to the generator monomials $x^a y^b z^c$, $a + b + c = n$, gives a more usable form of the basic sets similar to that given for the wave equation in (14) with $d = 0$ and $t^{2j}/(2j)!$ replaced by $t^{2j}/(1+k)(3+k)\ldots(2j-1+k)\cdot 2^j j!$.

However, if k is negative, the system of differential equations (4) associated with the solution of (18) under conditions (2) has a solution,

$$(19) \quad u_j = \sum_{n=0}^{j} \frac{a_n t^{2j-2n+1-k}}{2^{j-n}(j-n)!\,(1-k)(3-k)\ldots(2j-2n+1-k)} + \frac{t^{2j}}{2^j j!(1+k)(3+k)\ldots(2j-1+k)},$$

provided $k \neq -1, -3, \ldots, -(2j-1)$, which is not unique since the a_n are arbitrary.

Weinstein has recently shown (see for instance **(3)**), that for odd negative integral values of k, solutions of (18) satisfying certain differentiability conditions exist only if the initial value function is polyharmonic of order $(1-k)/2$ and that, in this case, the addition to the solution of any function of the type

$$t^{1-k}\, u^{2-k}(x_1, \ldots x_n, t),$$

(u^{2-k} denotes a solution of (18) with k replaced by $2-k$) which vanishes with its t derivative at $t = 0$, gives another solution of the problem. We may illustrate this result of Weinstein by applying Theorem 1 to (18) with boundary conditions (16). As a first solution we may take $a_n = 0$, $n = 0, 1, 2, \ldots$ and obtain

$$(20)\quad u(x, y, z, t) = e^x \cos y \left[xz^4 + (z^4 + 6xz^2)\frac{t^2}{1+k} + (6z^2 + 3x)\frac{t^4}{(1+k)(3+k)} + \frac{3t^6}{(1+k)(3+k)(5+k)} \right].$$

The solution (20) is invalid for $k = -1, -3, -5$, but holds for $k = -7, -9, \ldots$ etc. Since $xe^x \cos y \cdot z^4$ is polyharmonic of order $\frac{1}{2}\{1-(-7)\} = 4$, this illustrates the result of Weinstein quoted above. As a further illustration of Weinstein's result we may let $a_0 \neq 0$ in (19) and augment the solution (20) by

$$(21)\quad a_0 \frac{t^{1-k}}{1-k} e^x \cos y \left[xz^4 + (z^4 + 6xz^2)\frac{t^2}{[1+(2-k)]} + (6z^2 + 3x)\frac{t^4}{[1+(2-k)][3+(2-k)]} + \frac{3t^6}{[1+(2-k)][3+(2-k)][5+(2-k)]} \right]$$

which is clearly an arbitrary constant multiple of t^{1-k} by a solution of (18) with k replaced by $(2-k)$ which vanishes with its t derivative at $t = 0$ for all negative k.

Where condition (3) does not hold, the above methods may lead us to solutions in infinite series form. As an illustration we may consider the vibrating string problem

$$(22)\quad u_{xx} - a^{-2} u_{tt} = 0,\ u(x, 0) = P(x),\ u_t(x, 0) = 0,\ u(0, t) = u(L, t) = 0,$$

where the function $P(x)$ has the Fourier expansion

$$(23)\qquad P(x) = \sum_1^\infty b_n \sin \frac{n\pi x}{L}.$$

We then obtain

$$\Phi^j P = \sum_{n=1}^{\infty} (-1)^j b_n \left(\frac{n\pi}{L}\right)^{2j} \sin \frac{n\pi x}{L} \tag{24}$$

and

$$u_j = \frac{(at)^{2j}}{(2j)!}, \qquad j = 0, 1, 2, \ldots . \tag{25}$$

Taking

$$u = \sum_{j=0}^{\infty} \sum_{n=1}^{\infty} (-1)^j b_n \left(\frac{n\pi}{L}\right)^{2j} \sin \frac{n\pi x}{L} \cdot \frac{(at)^{2j}}{(2j)!} \tag{26}$$

we have a formal solution. If the order of summation be interchanged we obtain

$$u = \sum_{n=1}^{\infty} b_n \sin \frac{n\pi x}{L} \cos \frac{n\pi at}{L}, \tag{27}$$

which is the usual form of the solution of this problem. The one dimensional heat flow problem may also be solved by this method but the applicability of the method to other problems, because of convergence questions, is a matter requiring further study.

References

1. E. P. Miles, Jr. and Ernest Williams, *A basic set of homogeneous harmonic polynomials in k variables*, Proc. Amer. Math. Soc., 6 (1955), 191–194.
2. E. P. Miles, Jr. and Ernest Williams, *A basic set of polynomial solutions for the Euler-Poisson-Darboux and Beltrami equations*. To appear in the American Mathematical Monthly.
3. Alexander Weinstein, *On the wave equation and the equation of Euler-Poisson*, Proc. Fifth Symposium in Applied Mathematics (New York, 1954), 137–147.

Alabama Polytechnic Institute,
Auburn, Alabama

IMPLICATIONS OF HADAMARD'S CONDITIONS FOR ELASTIC STABILITY WITH RESPECT TO UNIQUENESS THEOREMS

J. L. ERICKSEN AND R. A. TOUPIN

Introduction. The purpose of this paper is to discuss implications of Hadamard's condition for elastic stability **(2**, §269**)** with respect to uniqueness of solutions of boundary value problems in the theory of small deformations superimposed on large. We show that a slightly refined form of his condition implies a uniqueness theorem for displacement boundary value problems. We construct a counter-example showing that his condition does not imply uniqueness of solutions for one type of stress boundary value problem. Hadamard **(2**, Ch. VI**)** showed that his condition implies the reality of all possible velocities of propagation of acceleration waves. To our knowledge, this is the only other known consequence of his condition.

Truesdell **(8)** has focused attention on the question of what conditions should be imposed on the strain energy to exclude physically unacceptable behavior. We are indebted to him for discussing this problem with us, thereby stimulating our interest in the topics considered here, and for his constructive criticisms of our work.

It is sufficient for our purposes to require that all vector fields considered be of class C^2 at all points of the undeformed body, which points constitute a regular region of space $\mathfrak{R}$, as defined in **(3)**.

1. Elasticity theory. The theory of elasticity with which we are concerned is based on the existence of a strain energy per unit of undeformed volume Σ, which is a function of displacement gradients $U^\alpha{}_{,\beta}$. Here U^α are the components of the displacement vector referred to a material[1] coordinate system and the comma denotes covariant differentiation with respect to these coordinates. We assume Σ is of class C^3 for all $U^\alpha{}_{,\beta}$, that there are no constraints on the deformation, and that inertial and body forces vanish. The basic equations may then be written

$$(\partial\Sigma/\partial U^\alpha{}_{,\beta})_{,\beta} = 0, \tag{1}$$

as was shown by Kirchhoff **(4)**.

To obtain the equations of the theory of small deformations superimposed on (possibly) large deformations, one writes $U^\alpha = V^\alpha + W^\alpha$, linearizes Eq. (1) with respect to the W^α, and assumes that the displacement V^α satisfies (1).

Received September 13, 1955.

[1]The adjective "Lagrangian" is used more frequently. For reasons pointed out in **(7**, §14**)**, "material" is preferable.

We thus have

$$(2)\qquad \begin{cases} \partial\Sigma/\partial U^{\alpha}{}_{,\beta} = \partial\Sigma/\partial V^{\alpha}{}_{,\beta} + W^{\gamma}{}_{,\delta}\partial^2\Sigma/\partial V^{\alpha}{}_{,\beta}\partial V^{\gamma}{}_{,\delta}, \\ (\partial\Sigma/\partial V^{\alpha}{}_{,\beta})_{,\beta} = 0, \\ (W^{\gamma}{}_{,\delta}\partial^2\Sigma/\partial V^{\alpha}{}_{,\beta}\partial V^{\gamma}{}_{,\delta})_{,\beta} = 0. \end{cases}$$

Alternative formulations are given in **(1)** and **(6**, §55**)**. Here V^{α} is regarded as known, whereas W^{α} is to be determined by the linear equations (2) and appropriate boundary conditions. A displacement boundary value problem is set by specifying W^{α} on the bounding surface $\mathfrak{S}$ of $\mathfrak{R}$. To show that two displacement vectors satisfying these same conditions are identical, it suffices, because of linearity, to show that any W^{α} which satisfies (2) and vanishes on $\mathfrak{S}$ must vanish in $\mathfrak{R}$. For equations (1), one type of stress boundary value problem is set by specifying $N_{\beta}\partial\Sigma/\partial U^{\alpha}{}_{,\beta}$ on $\mathfrak{S}$, N_{α} being a unit vector normal to $\mathfrak{S}$. In the above linearized theory, this leads to a problem in which the quantities

$$(3)\qquad T_{\alpha} = N_{\beta}W^{\gamma}{}_{,\delta}\partial^2\Sigma/\partial V^{\alpha}{}_{,\beta}\partial V^{\gamma}{}_{,\delta}$$

are specified on $\mathfrak{S}$.

If one sets $V^{\alpha}{}_{,\beta} = 0$ in (2), W^{α} becomes a small deformation about the state of zero deformation. It is customary to assume that $\partial\Sigma/\partial V^{\alpha}{}_{,\beta} = 0$ when $V^{\alpha}{}_{,\beta} = 0$. We make no use of this assumption. If one places certain restrictions on $\partial^2\Sigma/\partial V^{\alpha}{}_{,\beta}\,\partial V^{\gamma}{}_{,\delta}$ evaluated at $V^{\alpha}{}_{,\beta} = 0$, one obtains the usual equations of the classical linear theory of elasticity. For example, for isotropic materials, one takes

$$(4)\qquad \partial^2\Sigma/\partial V^{\alpha}{}_{,\beta}\partial V^{\gamma}{}_{,\delta}|_{V^{\rho}{}_{,\sigma}=0} = \lambda\delta^{\beta}_{\alpha}\delta^{\delta}_{\gamma} + \mu(\delta^{\beta}_{\gamma}\delta^{\delta}_{\alpha} + G^{\beta\delta}G_{\alpha\gamma})$$

where λ and μ are the Lamé constants and $G_{\alpha\beta}$ is the metric tensor. The boundary data (3), which become

$$(5)\qquad T_{\alpha} = \lambda W^{\beta}{}_{,\beta}N_{\alpha} + \mu(W_{\alpha,\beta} + W_{\beta,\alpha})\,N^{\beta}$$

when (4) holds, is the data ordinarily prescribed in stress boundary value problems in the linear theory. The Kirchhoff uniqueness proof, valid when $(3\lambda + 2\mu)\,\mu > 0$, establishes that $T_{\alpha} = 0$ on $\mathfrak{S}$ implies that $W_{\alpha,\beta} + W_{\beta,\alpha} = 0$ in $\mathfrak{R}$. In other words, the boundary data (5) determines the displacement field W^{α} to within an infinitesimal rigid motion. It seems reasonable to expect that this uniqueness theorem will hold for the small deformation when $V^{\alpha}{}_{,\beta} \neq 0$ if suitable restrictions are placed on Σ and V^{α}. What constitutes a set of "suitable restrictions" on Σ is, according to Truesdell **(8)**, the main open problem in the theory of finite elastic deformations. We shall show that the desired uniqueness does not follow from Hadamard's stability condition.

2. Elastic stability. Hadamard **(2**, §269**)** calls a deformation stable whenever the second variation in total strain energy is non-negative for all variations in U^{α} which vanish on $\mathfrak{S}$. Formally, stability means that

(6) $$\Phi = \delta^2 \int_{\mathfrak{R}} \Sigma dV > 0$$

whenever $\delta U^\alpha = \delta^2 U^\alpha = 0$ on $\mathfrak{S}$, dV being the volume element. We have $\Phi = \Phi_1 + \Phi_2$, where

(7) $$\Phi_1 = \int_{\mathfrak{R}} \delta^2 U^\alpha{}_{,\beta} \partial\Sigma/\partial U^\alpha{}_{,\beta} dV,$$
$$\Phi_2 = \int_{\mathfrak{R}} \delta U^\alpha{}_{,\beta} \delta U^\gamma{}_{,\delta} \partial^2\Sigma/\partial U^\alpha{}_{,\beta} \partial U^\gamma{}_{,\delta} dV.$$

From (1), (6) and (7),

$$\Phi_1 = \int_{\mathfrak{R}} (\delta^2 U^\alpha \partial\Sigma/\partial U^\alpha{}_{,\beta})_{,\beta} dV = \oint_{\mathfrak{S}} \delta^2 U^\alpha \partial\Sigma/\partial U^\alpha{}_{,\beta} dS_\beta = 0$$

where dS_β is the vector element of area. Similarly, from (6) and (7),

(8) $$\Phi_2 = -\int_{\mathfrak{R}} \delta U^\alpha (\delta U^\gamma{}_{,\delta} \partial^2\Sigma/\partial U^\alpha{}_{,\beta} \partial U^\gamma{}_{,\delta})_{,\beta} dV.$$

Thus (6) can be replaced by

$$\Phi_2 > 0 \text{ wherever } \delta U^\alpha = 0 \text{ on } \mathfrak{S},$$

Φ_2 being given by (7) or (8). An analysis made by Kelvin **(5)** suggests that it is desirable to distinguish neutral or labile stability, for which $\Phi_2 = 0$ for some $\delta U^\alpha \not\equiv 0$, from ordinary stability, for which $\Phi_2 = 0$ implies $\delta U^\alpha \equiv 0$, and we find it essential for our purposes to make this distinction. Henceforth, "stability" means ordinary stability, neutral stability being excluded. There was no reason for Hadamard to make this distinction since the results which he obtained are insensitive to it.

3. Uniqueness. We begin by proving a uniqueness theorem for displacement boundary value problems.

THEOREM 1. *In the theory of small deformations superimposed on large, if the large deformation is stable, the displacement boundary value problem for the small deformation has at most one solution.*

Proof. Let W^α be any solution of (2) such that $W^\alpha = 0$ on $\mathfrak{S}$. Multiplying the last of equations (2) by W^α, summing on α, and integrating the result over $\mathfrak{R}$, we obtain

$$\int_{\mathfrak{R}} W^\alpha (W^\gamma{}_{,\delta} \partial^2\Sigma/\partial V^\alpha{}_{,\beta} \partial V^\gamma{}_{,\delta})_{,\beta} dV = 0.$$

From this and (8), we see that Φ_2, evaluated for $U^\alpha = V^\alpha$ and $\delta U^\alpha = W^\alpha$, vanishes. If V^α is stable, $\Phi_2 = 0$ implies $W^\alpha \equiv 0$. Thus, if a solution exists for a given displacement boundary value problem, it is unique.

We now proceed to determine necessary and sufficient conditions for the stability of the state of zero deformation of isotropic materials. In this case (4) holds and we obtain from (7) with $U^\alpha{}_{,\beta} = 0$,

$$\Phi_2 = \int_{\mathfrak{R}} [\lambda(\delta U^\alpha{}_{,\alpha})^2 + \mu(\delta U^\alpha{}_{,\beta}\delta U^\beta{}_{,\alpha} + \delta U^{\alpha,\beta}\delta U_{\alpha,\beta})]dV.$$

Using the fact that $\delta U^\alpha = 0$ on $\mathfrak{S}$, we have

$$0 = \int_{\mathfrak{S}} [\delta U^\alpha \delta U^\beta{}_{,\beta} - \delta U^\beta \delta U^\alpha{}_{,\beta}]dS_\alpha = \int_{\mathfrak{R}} [\delta U^\alpha \delta U^\beta{}_{,\beta} - \delta U^\beta \delta U^\alpha{}_{,\beta}]_{,\alpha} dV$$

$$= \oint_{\mathfrak{R}} [(\delta U^\alpha{}_{,\alpha})^2 - \delta U^\beta{}_{,\alpha}\delta U^\alpha{}_{,\beta}]dV.$$

Also,

$$\delta U^{\alpha,\beta}\delta U_{\alpha,\beta} = \delta U^\alpha{}_{,\beta}\delta U^\beta{}_{,\alpha} + 2\omega^{\alpha\beta}\omega_{\alpha\beta},$$

where $2\omega_{\alpha,\beta} = U_{\alpha,\beta} - U_{\beta,\alpha}$. Using these relations, we obtain

$$\Phi_2 = (\lambda + 2\mu) \int_{\mathfrak{R}} (\delta U^\alpha{}_{,\alpha})^2 dV + 2\mu \int_{\mathfrak{R}} \omega^{\alpha\beta}\omega_{\alpha\beta} dV, \tag{9}$$

a result due to Kelvin **(5)**. Since each integral is non-negative, we have stability, or at least neutral stability, of zero deformation, so long as $\lambda + 2\mu \geqslant 0$ and $\mu \geqslant 0$. A slightly sharper result is easily obtained.

LEMMA. *For stability of the state of zero deformation of an isotropic elastic material, it is necessary and sufficient that* $\lambda + 2\mu > 0$ *and* $\mu > 0$.

Proof of sufficiency. If $\lambda + 2\mu > 0$ and $\mu > 0$, it is clear from (9) that $\Phi_2 \geqslant 0$, the equality holding if and only if $\delta U^\alpha{}_{,\alpha} = \omega^\alpha{}_\beta = 0$. These conditions imply that $\delta U_\alpha = \phi_{,\alpha}$, where ϕ is harmonic. Since $\phi_{,\alpha} = 0$ on $\mathfrak{S}$ and ϕ is harmonic in $\mathfrak{R}$, $\phi_{,\alpha} \equiv 0$. Hence $\Phi_2 > 0$, unless $\delta U^\alpha \equiv 0$.

Proof of necessity. To show that $\Phi_2 > 0$ implies $\lambda + 2\mu > 0$, it suffices to construct functions δU^α such that $\delta U^\alpha = 0$ on $\mathfrak{S}$, $\omega^\alpha{}_\beta \equiv 0$ in $\mathfrak{R}$, $\delta U^\alpha{}_{,\alpha} \not\equiv 0$, as is clear from (9). One can take $\delta U_\alpha = \psi_{,\alpha}$, where ψ is any function, not a constant, whose gradient vanishes on $\mathfrak{S}$. For example if, $\mathfrak{C} \subset \mathfrak{R}$ is a sphere of radius $r_0 > 0$, we may take $\psi = 0$ in $\mathfrak{R} - \mathfrak{C}$, $\psi = (r - r_0)^4$ in $\mathfrak{C}$, where r denotes the distance measured from the center of $\mathfrak{C}$. Similarly, to show that $\Phi_2 > 0$ implies $\mu > 0$, one need only construct δU^α with $\delta U^\alpha = 0$ on $\mathfrak{S}$, $\delta U^\alpha{}_{,\alpha} \equiv 0$, $\delta U^\alpha \not\equiv 0$, and use (9). Such variations are easily constructed.

THEOREM 2. *For the stress boundary value problem* (3), *stability of the deformation* V^α *does not imply that the displacements* W^α *will be determined to within an infinitesimal rigid motion.*

Proof. It suffices to establish that uniqueness does not follow from stability in the special case when $V^\alpha{}_{,\beta} = 0$ and (4) holds. By the lemma, we have stability if $3\lambda + 2\mu = 0$, $\mu > 0$. It follows, using (4), that (2) is satisfied by any W^α such that $W^\alpha{}_{,\beta} = a\,\delta^\alpha{}_\beta$, where a is an arbitrary constant. For any

such displacement, (5), with $3\lambda + 2\mu = 0$, gives $T_a = 0$ for arbitrary N_a. If $a \neq 0$, the displacement considered above is not an infinitesimal rigid motion, whence the theorem follows.

This theorem indicates that the definition of stability used here leads to results in disagreement with the intuitive notion, expounded by many writers in stability, that such non-uniqueness should be associated with instability. This might be regarded as an indication that it would be desirable to introduce further criteria to enable one to refine further the classification of types of stability used here.

THEOREM 3. *Neutral stability of the deformation V^a does not imply uniqueness of solutions to displacement boundary value problems in the theory of small deformations superimposed on large.*

Proof. Again it suffices to establish the theorem in the special case when (4) holds with $\lambda + 2\mu = 0$, $\mu > 0$. From (9) and the lemma, we then have neutral stability, but not stability. From the proof of the lemma, we can construct functions $W^a = \psi^{,a}$ such that $W^a = 0$ on $\mathfrak{S}$, $W^a \not\equiv 0$ in $\mathfrak{R}$. It follows easily, using (4), that when $\lambda + 2\mu = 0$, any such displacement satisfies (2). Since $W^a \equiv 0$ is another solution satisfying the same boundary conditions, we do not have uniqueness.

Theorems 1 and 3 illustrate the importance of distinguishing between ordinary and neutral stability. As is pointed out by Whittaker (9, pp. 145–148), the case $\lambda + 2\mu = 0$, $\mu > 0$ is of some historical interest, having been considered as an aether theory.

REFERENCES

1. A. E. Green, R. S. Rivlin, and R. T. Shield, *General theory of small elastic deformations superposed on finite elastic deformations*, Proc. Roy. Soc. London (A), *211* (1951), 128–154.
2. J. Hadamard, *Leçons sur la propagation des ondes et les équations de l'hydrodynamique* (Paris, 1903).
3. O. D. Kellogg, *Foundations of Potential Theory* (New York, 1929).
4. G. Kirchhoff, *Ueber die Gleichungen des Gleichgewichts eines elastischen Körpers bei nicht unendlich kleinen Verschiebungen seiner Theile*, Akad. Wiss. Wien Sitz., *9* (1852), 762–773.
5. W. Thomson (Lord Kelvin), *On the reflexion and refraction of light*, Phil. Mag., *26* (1888), 414–425.
6. C. Truesdell, *The mechanical foundations of elasticity and fluid dynamics*, J. Rational Mech. Anal., *1* (1952), 125–300.
7. C. Truesdell, *The Kinematics of Vorticity* (Bloomington, 1954).
8. C. Truesdell, *Das ungelöste Hauptproblem der endlichen Elastizitätstheorie*, to appear in Z.a.M.M.
9. E. Whittaker, *A History of the Theories of Aether and Electricity* (New York, 1951).

Applied Mathematics Branch,
Naval Research Laboratory, Washington, D.C.

A GENERALIZED AVERAGING OPERATOR

D. B. SUMNER

1. Introduction. The averaging operator $\nabla f(z) = \frac{1}{2}[f(z+h) + f(z)]$ has an extensive literature, the most detailed account being that of Nörlund **(4)**. In discussing solutions of the functional relation

$$\nabla f(z) = \phi(z), \tag{1.1}$$

he defines a "principal solution" **(4**, p. 41**)** by means of a summability process, and later, working in terms of complex numbers, he obtains **(4**, p. 70**)** a principal solution of (1.1) by means of a contour integral. He distinguishes his principal solution from other solutions, by showing that it is continuous at $h = 0$. His work includes a detailed account of the polynomial solutions of

$$\nabla f(z) = z^k, \tag{1.2}$$

the Euler polynomials with assigned values at $z = \frac{1}{2}$. Milne-Thomson **(3**, pp. 519–521**)** gives an account of generalized Euler numbers arising from the operator ∇^N, (N a positive integer) and of the generalized Euler numbers.

In this paper the ideas of Milne-Thomson are taken a step further. The operator ∇^λ is defined for all real λ, and is shown to be applicable to a wide class of functions. Polynomials corresponding to the generalized Euler polynomials of Milne-Thomson and a sequence of numbers corresponding to Nörlund's C-numbers **(4**, p. 27**)** are defined and some of their more important properties established. The inverse operator $\nabla^{-\lambda}$ is defined, and is shown to invert the operation ∇^λ and to give a unique solution in terms of the functions to which ∇^λ is applicable.

2. Generalized power of the averaging operator. The averaging (or mean) operator is defined for span h by

$$\nabla f(z) = \tfrac{1}{2}[f(z+h) + f(z)], \tag{2.1}$$

and its positive integer powers by

$$\nabla^M f(z) = \nabla\nabla^{M-1} f(z) = \sum_0^M \binom{M}{p} f(z+hp)/2^M. \tag{2.2}$$

To define $\nabla^\lambda f(z)$, where λ is related to the positive integer N by

$$N - 1 < \lambda < N, \tag{2.3}$$

we use the formal relation

Received November 7, 1955.

$$\nabla f(z) = \tfrac{1}{2}(1 + \exp hD).f(z),$$

and write

$$\nabla^\lambda = \frac{(1 + \exp hD)^{N+1}}{2^\lambda(1 + \exp hD)^\mu}, \qquad \mu = N + 1 - \lambda.$$

The operation in the numerator can be expressed by means of (2.2); and to obtain a representation of the operation in the denominator, we use the fact that

$$\frac{1}{(1 + \exp t)^\alpha} = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{\exp(-tw)dw}{E(\alpha, w)},$$

where t is real, α is positive, $0 < c < \alpha$ and

$$E(\alpha, w) = \Gamma(\alpha)/\Gamma(w)\,\Gamma(\alpha - w). \tag{2.4}$$

Using the abbreviation

$$\int_c \text{ for } \int_{c-i\infty}^{c+i\infty},$$

we then have formally

$$\begin{aligned} \nabla^\lambda f(z) &= \frac{1}{2^\lambda}\sum_0^{N+1}\binom{N+1}{p} e^{phD}\cdot \int_c \frac{\exp(-hDw)dw}{2\pi i\,E(\mu, w)}\cdot f(z) \\ &= \sum_0^{N+1}\binom{N+1}{p}\int_c \frac{f(z + ph - hw)dw}{2\pi i\,E(\mu, w)\,2^\lambda}, \end{aligned} \tag{2.5}$$

on using the shift operation $\exp(kD).f(z) = f(z + k)$. We take (2.5) *as the definition* of $\nabla^\lambda f(z)$, *if* λ *satisfies* (2.3), *the span* h *is positive or negative and the integrals exist.*

Although less restrictive assumptions as to the nature of $f(z)$ would be sufficient to ensure the existence of the integrals in (2.5), we shall assume throughout that

(2.6) *$f(z)$ is an entire function of exponential order κ, $\kappa h < \pi$.*

The following proposition is then an easy consequence of (2.6) and the fact that

$$|\Gamma(c + iv)\,\Gamma(\mu - c - iv)| \sim A\exp(-\pi|v|)\,.\,|v|^{N-\lambda}, \quad (|v| \to \infty):$$

if $\phi(z, h)$ is the function defined by (2.5) *and $f(z)$ satisfies* (2.6), *then $\phi(z, h)$ is an entire function of exponential order κ (in z) and*

$$\lim_{h\to 0}\phi(z, h) = f(z). \tag{2.7}$$

Thus $\phi(z, h)$ has the property (2.7) which was noted by Nörlund (**4**, p. 46) as being characteristic of his principal solution of the functional equation $\nabla f(z) = \phi(z)$. It must be observed, however, that there do exist entire functions

in z, for example, $\cos(\pi z/h)$ which satisfy neither (2.6) nor (2.7), but for which the operation ∇^λ is defined when λ is a positive integer but not otherwise.

In the particular case when $\lambda = N$, the definition (2.5) gives for $f(z)$ satisfying (2.6),

$$\begin{aligned}\nabla^N f(z) &= 2^{-N}\sum_0^N \binom{N}{p} \int_c \frac{f[z+h(p-w)] + f[z+h(p+1-w)]}{2\pi i\, E(1,w)}\, dw \\ &= 2^{-N}\sum_0^N \binom{N}{p} \operatorname{Res}\left\{\frac{\pi f[z+h(p-w)]}{\sin \pi w}; 0\right\} \\ &= 2^{-N}\sum_0^N \binom{N}{p} f(z+ph),\end{aligned}$$

which is the value given in (2.2).

We may confine ourselves to cases where $h > 0$ by reason of the following *extension property*: if $\phi(z, h) = \nabla^\lambda f(z)$, then

$$\phi(z + h\lambda, -h) = \phi(z, h). \tag{2.8}$$

For reversing the summation, and making the change of variable $w = \mu - \xi$, we have

$$\begin{aligned}\phi(z+h\lambda, -h) &= 2^{-\lambda}\sum_0^{N+1} \binom{N+1}{q} \int_{\mu-c} \frac{f(z+hq-h\xi)d\xi}{2\pi i\, E(\mu, \mu-\xi)} \\ &= 2^{-\lambda}\sum_0^{N+1} \binom{N+1}{q} \int_c \frac{f(z+hq-h\xi)d\xi}{2\pi i\, E(\mu,\xi)},\end{aligned}$$

by Cauchy's theorem, since $0 < c < \mu$, $0 < \mu - c < \mu$, and $E(\mu, \mu - \xi) = E(\mu, \xi)$.

3. The exponential property of ∇^λ.

We prove that

$$\nabla^\alpha \nabla^\beta f(z) = \nabla^{\alpha+\beta} f(z) \tag{3.1}$$

when α, β are positive. On account of (2.2) it is sufficient to give details for the cases

$$0 < \alpha + \beta < 1, \tag{3.2}$$

$$1 < \alpha + \beta < 2. \tag{3.3}$$

For the proof in the case (3.2) write $\alpha + \beta = \gamma$. Then

$$\nabla^\gamma f(z) = \sum_{n=0}^2 \binom{2}{n} \int_c \frac{f(z+hn-hw)\, dw}{2^\gamma 2\pi i\, E(2-\gamma, w)} \qquad (0 < c < 2 - \gamma); \tag{3.4}$$

and for $0 < a < 2 - \alpha$, $0 < b < 2 - \beta$,

$$\begin{aligned}\nabla^\alpha\nabla^\beta f(z) &= \sum_{p,q=0}^2 \binom{2}{p}\binom{2}{q} \int_a \frac{ds}{2^\gamma 2\pi i\, E(2-\alpha, s)} \int_b \frac{f[z+h(p+q-s-w)]\, dw}{2\pi i\, E(2-\beta, w)} \\ &= \sum_{n=0}^2 \binom{2}{n} \int_a \frac{ds}{2^\gamma 2\pi i\, E(2-\alpha,s)} \int_b \frac{F(s+w)\, dw}{2\pi i\, E(2-\beta, w)},\end{aligned} \tag{3.5}$$

where $F(\xi) = f[z + h(n - \xi)] + 2f[z + h(n + 1 - \xi)] + f[z + h(n + 2 - \xi)]$. By Cauchy's theorem we may take $0 < a < b$; then

$$\int_b \frac{F(s + w)\, dw}{2\pi i\, E(2 - \beta, w)} = \int_b \frac{F(\xi)\, d\xi}{2\pi i\, E(2 - \beta, \xi - s)}.$$

Hypothesis (2.6) guarantees the absolute convergence of the integrals in (3.5), so that

$$\begin{aligned}\nabla^\alpha \nabla^\beta f(z) &= \sum_0^2 \binom{2}{n} \int_b \frac{F(\xi)\, d\xi}{2\pi i} \int_a \frac{\Gamma(s)\,\Gamma(2 - \beta - \xi + s)\,\Gamma(2 - \alpha - s)\,\Gamma(\xi - s)\, ds}{2^\gamma 2\pi i\, \Gamma(2 - \alpha)\,\Gamma(2 - \beta)} \\ &= \sum_0^2 \binom{2}{n} \int_b \frac{\Gamma(\xi)\,\Gamma(4 - \gamma - \xi)\, F(\xi)\, d\xi}{2^\gamma 2\pi i\, \Gamma(4 - \gamma)},\end{aligned}$$

by Barnes's Lemma (**1**, p. 155). Abbreviating this expression as

$$2^{-\gamma} \sum_0^2 \binom{2}{n} [I_1 + 2I_2 + I_3]$$

we let the lines of integration in I_2 and I_3 be changed to $b + 1$ and $b + 2$ respectively; and since the only positive poles of the integrand are at $\xi = 4 - \gamma, 5 - \gamma, \ldots$ and since $4 - \gamma > 3$, no poles lie in the strip $b < R(\xi) < b + 2$. Cauchy's theorem may then be applied to give

$$\begin{aligned}I_1 + 2I_2 + I_3 &= \int_b \frac{[\Gamma(\xi)\Gamma(4-\gamma-\xi) + 2\Gamma(\xi+1)\Gamma(3-\gamma-\xi) + \Gamma(\xi+2)\Gamma(2-\gamma-\xi)] f[z + h(n-\xi)]\, d\xi}{2\pi i\, \Gamma(4 - \gamma)} \\ &= \int_b \frac{\Gamma(\xi)\,\Gamma(2 - \gamma - \xi)\, f[z + h(n - \xi)]\, d\xi}{2\pi i\, \Gamma(2 - \gamma)}.\end{aligned}$$

Thus we have from (3.4)

$$\nabla^\alpha \nabla^\beta f(z) = \nabla^{\alpha+\beta} f(z).$$

In the case (3.3)

$$\begin{aligned}\nabla^\gamma f(z) &= \sum_0^3 \binom{3}{n} \int_c \frac{f[z + h(n - w)]\, dw}{2^\gamma 2\pi i\, E(3 - \gamma, w)}, \\ \nabla^\alpha \nabla^\beta f(z) &= \sum_{p,q=0}^2 \binom{2}{p}\binom{2}{q} \int_a \frac{ds}{E(2 - \alpha, s)} \int_b \frac{f[z + h(p + q - s - w)]\, dw}{2^\gamma (2\pi i)^2\, E(2 - \beta, w)} \\ &= \sum_0^3 \binom{3}{n} \int_a \frac{ds}{2\pi i\, E(2 - \alpha, s)} \int_b \frac{\{f[z + h(n - s - w)] + f[z + h(n + 1 - s - w)]\}\, dw}{2\pi i\, E(2 - \beta, w)},\end{aligned}$$

and the previous argument may then be used to establish the result.

4. The numbers g_k^λ and the polynomials $g_k^\lambda(z)$. We digress here to define certain fundamental numbers and polynomials associated with ∇^λ. Let

$$\frac{2^\lambda}{(1+\exp t)^\lambda} = \sum_0^\infty \frac{g_k^\lambda t^k}{k!} \qquad (|t| < \pi), \tag{4.1}$$

$$g_k^\lambda(z) = \sum_0^k \binom{k}{m} z^{k-m} g_m^\lambda. \tag{4.2}$$

On writing $G(t) = 2^\lambda(1+\exp t)^{-\lambda}$, we obtain

$$[1+\exp(-t)]\,G'(t) + \lambda G(t) = 0,$$

$$\sum_{k=0}^n \binom{n}{k} [1+\exp(-t)]^{(k)} G^{(n+1-k)}(t) + \lambda G^{(n)}(t) = 0,$$

from which, on setting $t = 0$, and using the definition $g_k^\lambda = G^{(k)}(0)$, we have the recurrence relations

$$\sum_1^n \binom{n}{k} (-)^k g_{n+1-k}^\lambda + 2g_{n+1}^\lambda + \lambda g_n^\lambda = 0$$

or

$$(-)^n \sum_0^{n-1} \binom{n}{p} (-)^p g_{p+1}^\lambda + 2g_{n+1}^\lambda + \lambda g_n^\lambda = 0, \qquad n > 0, \tag{4.3}$$

$$g_0^\lambda = 1.$$

It is an easy calculation to establish for the polynomials $g_k^\lambda(z)$ the *generating relation*

$$\frac{2^\lambda \exp(zt)}{(1+\exp t)^\lambda} = \sum_0^\infty \frac{t^k g_k^\lambda(z)}{k!}. \tag{4.4}$$

The numbers g_k^λ have the following *explicit value* in terms of the Stirling numbers:

$$g_k^\lambda = (-)^k \sum_{p=1}^k \mathscr{S}_k^p\, \Gamma(\lambda+p)/\Gamma(\lambda)\, 2^p. \tag{4.5}$$

For we have

$$g_k^\lambda = \lim_{t\to 0} G^{(k)}(t) = \lim_{t\to 0} \frac{(-)^k 2^\lambda}{2\pi i} \int_c \frac{w^k \exp(-tw)\,dw}{E(\lambda, w)}$$
$$= \lim_{t\to 0} (-)^k 2^\lambda \sum_1^k \frac{\mathscr{S}_k^p \Gamma(\lambda+p)}{\Gamma(\lambda)\, 2\pi i} \int_c \frac{\exp(-tw)\,dw}{E(\lambda+p, \lambda-w)},$$

where

$$\mathscr{S}_k^p = \lim_{x\to 0} \frac{\Delta^p x^k}{p!}$$

are the Stirling numbers of the second kind (**2**, p. 134), and use is made of the identity

$$w^k = \sum_{p=1}^k \mathscr{S}_k^p\, \Gamma(w+p)/\Gamma(w).$$

Thus, using the notation $(\lambda)_p = \Gamma(\lambda + p)/\Gamma(\lambda)$,

$$g_k^\lambda = \lim_{t \to 0} (-)^k 2^\lambda \sum_{p=1}^{k} \frac{\mathscr{S}_k^p(\lambda)_p}{(1+e^t)^{\lambda+p}} = (-)^k 2^{-p} \sum_{p=1}^{k} \mathscr{S}_k^p(\lambda)_p.$$

We prove next that

$$\nabla^\lambda h^k g_k^\lambda(z/h) = z^k. \tag{4.6}$$

Writing $\xi = z/h$, we have from (4.4),

$$\frac{2^\lambda e^{t(\xi+n-w)}}{(1+e^t)^{\lambda-1}} = \sum_0^\infty \frac{t^k}{k!} [g_k^\lambda(\xi + n - w) + g_{k+1}^\lambda(\xi + n + 1 - w)].$$

Multiplying throughout by

$$\binom{N}{n} \Big/ 2\pi i\, 2^\lambda E(N + 1 - \lambda, w),$$

summing from $n = 0$ to N, and integrating with respect to w along the line $R(w) = c$ makes the right-hand side equal to

$$\sum_0^\infty \frac{t^k}{k!} \nabla^\lambda g_k^\lambda(\xi),$$

and the left-hand side equal to

$$\frac{e^{t\xi}}{(1+e^t)^{\lambda-1}} \sum_{n=0}^{N} \binom{N}{n} e^{nt} \int_c \frac{\exp(-tw)dw}{2\pi i\, E(N+1-\lambda, w)}$$
$$= \frac{e^{t\xi}(1+e^t)^N}{(1+e^t)^{\lambda-1}(1+e^t)^{N+1-\lambda}} = e^{t\xi}.$$

Thus

$$e^{t\xi} = \sum_0^\infty \frac{t^k}{k!} \nabla^\lambda g_k^\lambda(\xi),$$

and the result stated follows by comparing coefficients.

We note here that the function $h^k g_k^\lambda(z/h)$ has the property (2.7).

5. The inverse operator. A definition for negative powers of ∇ is obtained from the observation that formally

$$\begin{aligned} \nabla^{-\lambda}\phi(z) &= \frac{2^\lambda}{(1 + \exp hD)^\lambda}\phi(z) = \frac{2^\lambda}{2\pi i}\int_c \frac{\exp(-hDw)dw}{E(\lambda, w)} \cdot \phi(z) \\ &= \frac{2^\lambda}{2\pi i}\int_c \frac{\phi(z - hw)dw}{E(\lambda, w)}, \qquad 0 < c < \lambda. \end{aligned} \tag{5.1}$$

We take (5.1) *as the definition of* $\nabla^{-\lambda}\phi(z)$, and as before assume that $\phi(z)$ is of exponential order κ, $\kappa h < \pi$, as a sufficient condition for assuring the existence of the integral in (5.1). This definition is valid for any real h, but

we are justified in confining ourselves to the case $h > 0$ by the following *extension property*

(5.2) *if* $f(z) = \nabla^{-\lambda}\phi(z)$, *then* $f(z - h\lambda, -h) = f(z, h)$.

For on setting $w = \lambda - \xi$, and observing that

$$0 < \mathscr{R}(\xi) = \lambda - c < \lambda,$$

we may apply Cauchy's theorem to see that

$$\begin{aligned} 2^{-\lambda} f(z - h\lambda, -h) &= \int_c \frac{\phi(z - \lambda h + hw)dw}{2\pi i\, E(\lambda, w)} \qquad (0 < c < \lambda) \\ &= \int_{\lambda - c} \frac{\phi(z - h\xi)d\xi}{2\pi i\, E(\lambda, \lambda - \xi)} \\ &= \int_c \frac{\phi(z - h\xi)d\xi}{2\pi i\, E(\lambda, \xi)} = 2^{-\lambda} f(z, h). \end{aligned}$$

The definition (5.1) is easily applied in special cases. Since

$$\begin{aligned} (5.3)\quad \nabla^{-\lambda}\phi(z) &= \frac{2^\lambda}{2\pi i}\int_c \frac{dw}{E(\lambda, w)} \sum_{m=0}^{\infty} \frac{(-hw)^m}{m!}\,\phi^{(m)}(z) \\ &= \phi(z) + \frac{2^\lambda}{2\pi i}\int_c \frac{dw}{E(\lambda, w)} \sum_{m=1}^{\infty} \frac{(-h)^m}{m!}\,\phi^{(m)}(z) \sum_{p=1}^{m} \mathscr{S}_m^p \frac{\Gamma(w+p)}{\Gamma(w)} \\ &= \phi(z) + 2^\lambda \sum_{m=1}^{\infty} \frac{(-h)^m}{m!}\,\phi^{(m)}(z) \sum_{p=1}^{m} \mathscr{S}_m^p \frac{\Gamma(p+\lambda)}{\Gamma(\lambda)} \\ &\qquad \int_c \frac{\Gamma(w+p)\,\Gamma(\lambda - w)dw}{2\pi i\,\Gamma(p+\lambda)} \\ &= \phi(z) + \sum_{m=1}^{\infty} \frac{(-h)^m}{m!}\,\phi^{(m)}(z) \sum_{p=1}^{m} \mathscr{S}_m^p \frac{\Gamma(p+\lambda)}{\Gamma(\lambda)\,2^p} \\ &= \phi(z) + \sum_{m=1}^{\infty} \frac{h^m}{m!}\,\phi^{(m)}(z)\, g_m^\lambda = \sum_{m=0}^{\infty} \frac{h^m\,\phi^{(m)}(z)}{m!}\, g_m^\lambda, \end{aligned}$$

by (4.5) when the series converge. Thus when $\phi(z) = z^k$,

$$(5.4)\qquad \nabla^{-\lambda}\phi(z) = z^k + \sum_{1}^{k} \binom{k}{m} h^m z^{k-m} g_m^\lambda \cdot = \sum_{0}^{k} \binom{k}{m} h^m z^{k-m} g_m^\lambda .$$

Other simple cases would be

$$(5.5)\qquad \nabla^{-\lambda} e^z = 2^\lambda e^z/(1 + e^h)^\lambda$$

$$(5.6)\qquad \nabla^{-\lambda} \sin z = \sin\left(z - \frac{h}{2}\right) \Big/ \left(\cos\frac{h}{2}\right)^\lambda.$$

That the operation (5.1) does indeed invert $\nabla^\lambda f(z)$ is shown in the theorem:

THEOREM. *If* $\phi(z) = O(\exp \kappa|z|)$, $(|z| \to \infty)$, $\kappa h < \pi$, *and* $F(z)$ *is defined by* (5.1), *then* $F(z)$ *is of exponential order* κ, *and*

$$(5.7)\qquad \nabla^\lambda F(z) = \phi(z).$$

That $F(z) = O(\exp \kappa|z|)$ may be proved in a manner similar to that by which (2.7) was established. To prove (5.7), let $0 < a < N + 1 - \lambda$, $0 < b < \lambda$, and consider

$$\nabla^\lambda F(z) = 2^{-\lambda} \sum_{p=0}^{N+1} \binom{N+1}{p} \int_a \frac{F(z+ph-hs)ds}{2\pi i\, E(N+1-\lambda, s)}$$

$$= \sum_0^{N+1} \binom{N+1}{p} \int_a \frac{ds}{2\pi i\, E(N+1-\lambda, s)} \int_b \frac{\phi[z+ph-h(s+w)]dw}{2\pi i\, E(\lambda, w)}$$

$$= \sum_0^{N+1} \binom{N+1}{p} \int_a \frac{ds}{2\pi i\, E(N+1-\lambda, s)} \int_{a+b} \frac{\phi(z+ph-h\xi)d\xi}{2\pi i\, E(\lambda, \xi - s)}.$$

Since $a + b < a + \lambda$, and the poles of the inner integrand lie on the lines

$$\mathscr{R}(\xi) = a, a-1, \ldots, \mathscr{R}(\xi) = a+\lambda, a+\lambda+1, \ldots.$$

Cauchy's theorem may be applied to give

$$\nabla^\lambda F(z) = \sum_0^{N+1} \binom{N+1}{p} \int_a \frac{ds}{2\pi i\, E(N+1-\lambda, s)} \int_b \frac{\phi(z+ph-h\xi)d\xi}{2\pi i\, E(\lambda, \xi - s)}.$$

The exponential order of $\phi(z)$ and the order properties on vertical lines of the Γ-function (**5**, p. 151), are sufficient to establish the absolute convergence of this iterated integral, and Fubini's theorem may be applied to give

$$\nabla^\lambda F(z) = \sum_0^{N+1} \binom{N+1}{p} \int_b \frac{\phi(z+ph-h\xi)d\xi}{2\pi i} \int_a \frac{\Gamma(s)\,\Gamma(\lambda-\xi+s)\,\Gamma(N+1-\lambda-s)\,\Gamma(\xi-s)ds}{2\pi i\,\Gamma(\lambda)\,\Gamma(N+1-\lambda)}$$

$$= \sum_0^{N+1} \binom{N+1}{p} \int_b \frac{\phi(z+ph-h\xi)d\xi}{2\pi i} \int_{L_a} \frac{\Gamma(s)\,\Gamma(\lambda-\xi+s)\,\Gamma(N+1-\lambda-s)\,\Gamma(\xi-s)ds}{2\pi i\,\Gamma(\lambda)\,\Gamma(N+1-\lambda)}$$

by Cauchy's theorem, where the contour L_a is obtained by deforming $R(s) = a$ in such a way that the poles of $\Gamma(N+1-\lambda-s)\,\Gamma(\xi-s)$ lie to the right of L_a, while the poles of $\Gamma(s)\,\Gamma(\lambda-\xi+s)$ lie to the left. Then by Barnes's Lemma (**1**, p. 155),

$$\nabla^\lambda F(z) = \sum_0^{N+1} \binom{N+1}{p} \int_b \frac{\phi[z-h(\xi-p)]d\xi}{2\pi i\, E(N+1, \xi)} = A + B.$$

To evaluate

$$A = \sum_0^{N} \binom{N+1}{p} \int_b \frac{\phi[z-h(\xi-p)]d\xi}{2\pi i\, E(N+1, \xi)},$$

Cauchy's theorem may be applied, since $0 < p < N$, to give

$$\begin{aligned} A &= \sum_0^N \binom{N+1}{p} \int_{b+p} \frac{\phi[z - h(\xi - p)]d\xi}{2\pi i\, E(N+1, \xi)} \\ &= \sum_0^N \binom{N+1}{p} \int_b \frac{\phi(z - h\xi)d\xi}{2\pi i\, E(N+1, p+\xi)} \\ &= \int_b \frac{\phi(z - h\xi)}{2\pi i} \sum_0^N \binom{N+1}{p} \frac{\Gamma(p+\xi)\, \Gamma(N+1-p-\xi)}{\Gamma(N+1)} d\xi \\ &= \int_b \frac{\phi(z - h\xi)}{2\pi i\, E(N+1, \xi)} \sum_0^N \frac{(-N-1)_p(\xi)_p}{p!\, (\xi - N)_p} d\xi \end{aligned}$$

where

$$\begin{aligned} \sum_0^N \frac{(-N-1)_p(\xi)_p}{p!\, (\xi - N)_p} &= {}_2F_1\left[\begin{matrix} -N-1, \xi \\ \xi - N \end{matrix}; 1\right] - \frac{(-)^{N+1}(\xi)_{N+1}}{(\xi - N)_{N+1}} \\ &= \frac{(-N)_{N+1}}{(\xi - N)_{N+1}} - \frac{\Gamma(N+1+\xi)\, \Gamma(-\xi)}{\Gamma(\xi)\, \Gamma(N+1-\xi)} \\ &= -\frac{\Gamma(N+1+\xi)(-\xi)}{\Gamma(\xi)\, \Gamma(N+1-\xi)}. \end{aligned}$$

Thus $A + B$

$$\begin{aligned} &= \int_b \frac{\Gamma(\xi)\, \Gamma(N+1-\xi)\, \phi[z-h(\xi-N-1)] - \Gamma(N+1+\xi)\, \Gamma(-\xi)\, \Gamma(z-h\xi)}{2\pi i\, \Gamma(N+1)} d\xi \\ &= \left\{ \int_{b-N-1} - \int_b \right\} \frac{\Gamma(N+1+w)\, \Gamma(-w)\, \phi(z - hw)dw}{2\pi i\, \Gamma(N+1)} \\ &= -\operatorname{Res}\left\{ \frac{\Gamma(N+1+w)\, \Gamma(-w)\, \phi(z - hw)}{\Gamma(N+1)}; 0 \right\} = \phi(z), \end{aligned}$$

which completes the proof.

6. Remarks. It is well known that the functional equation

$$\nabla^N f(z) = \phi(z), \qquad (N = 1, 2, \ldots) \tag{6.1}$$

has solutions other than that given by (5.1). For example, if $p(z)$ has the property

$$p(z + h) + p(z) = 0, \tag{6.2}$$

it is a solution of the homogeneous equation

$$\nabla^N f(z) = 0;$$

and if it is added to the solution of (6.1) given by (5.1), the resulting function is still a solution of (6.1). It does not, however, have the property (2.7), since for example $p(z)$ could be $\sin(\pi z/h)$ or $\cos(\pi z/h)$. Moreover it need not satisfy requirement (2.6), since

$$\cos(\pi z/h) = O[\exp(\pi|y|/h)], \qquad (|y| \to \infty),$$

and ∇^λ need not then be defined except when $\lambda = 1, 2, \ldots$. These facts suggest the possible existence of a set of eigenvalues $\lambda = 1, 2, \ldots$, with a family of eigenfunctions corresponding to each eigenvalue for the operator ∇^λ.

References

1. E. W. Barnes, *A new development of the theory of the hypergeometric functions*, Proc. London Math. Soc. (2), *6* (1908), 141–177.
2. C. Jordan, *Calculus of Finite Differences* (New York, 1947).
3. L. Milne-Thomson, *Two classes of generalized polynomials*, Proc. London Math. Soc. (2), *35* (1933), 514–522.
4. N. E. Nörlund, *Vorlesungen über Differenzenrechnung* (Berlin, 1924).
5. E. C. Tirchmarsh, *The Theory of Functions* (2nd. ed., Oxford 1939).

Hamilton College,
McMaster University

NOTE ON A STIELTJES TYPE OF INVERSION

PASQUALE PORCELLI

If $F(z)$ is an analytic function for $z \notin [-\infty, -1]$, $g(t)$ of bounded variation and real valued for $0 < t < 1$ and

$$F(z) = \int_0^1 (1 + zt)^{-1} dg(t),$$

then the Stieltjes type of inversion between $F(z)$ and $g(t)$ (cf. **1**, p. 339, Theorem 7a) is

$$\lim_{y \to 0+} \frac{-1}{\pi} \int_v^u \frac{1}{t} I_m F\left(-\frac{1}{t} + iy\right) dt = \frac{g(u+) + g(u-)}{2} - \frac{g(v+) + g(v-)}{2},$$

where $0 < v < u < 1$, $I_m F(z)$ is the imaginary part of $F(z)$ and $z = -t^{-1} + iy$.

A second type of inversion between $F(z)$ and $g(t)$ was obtained by Widder (**1**, p. 340, Theorem 7b) under the additional hypothesis that $g(t)$ is an absolutely continuous function. In the following theorem we shall establish an inversion between $F(z)$ and the right- and left-hand derivatives of $g(t)$ without the restriction that $g(t)$ be an integral.

THEOREM. *Let $F(z)$ be analytic for $z \notin [-\infty, -1]$, $g(t)$ real valued and of bounded variation on* $[0, 1]$ *and*

$$F(z) = \int_0^1 (1 + zt)^{-1} dg(t), \tag{1}$$

then

$$\lim_{y \to 0+} \frac{-1}{\pi t} I_m F\left(\frac{-1}{t} + iy\right) = \frac{g'^{+}(t) + g'^{-}(t)}{2}$$

for any t in $(0, 1)$ *at which the right- and left-hand derivatives $g'^{+}(t)$ and $g'^{-}(t)$ exist.*

Proof. Let us suppose that $g(0) = 0$, $0 < t_0 < 1$ and that $g'^{+}(t_0)$ and $g'^{-}(t_0)$ exist. If we set

$$R(t) = [(t_0 - t)^2 + (t_0 y t)^2]^{-1}$$

and $s = t_0 \pi^{-1}$, then from (1) we have

$$\frac{-1}{\pi t_0} I_m F\left(\frac{-1}{t} + iy\right) = sy \int_0^{t_0} t R(t) dg(t) + sy \int_{t_0}^1 t R(t) dg(t). \tag{2}$$

Received December 20, 1955 and in revised form March 9, 1956. Work on this paper was sponsored by the Air Research and Development Command, under Contract AF 18 (600) 1393.

In the first integral of this expression we can replace $g(t)$ by $[g(t_0) + g'^{-}(t_0)(t - t_0) + h(t)(t - t_0)]$, where $h(t)$ is continuous at t_0 and $h(t_0) = 0$, so that

$$sy\int_0^{t_0} tR(t)dg(t) = syg'^{-}(t_0)\int_0^{t_0} tR(t)dt + sy\int_0^{t_0} tR(t)dh(t)(t - t_0).$$

The first term on the right side of this equation can be integrated directly and we can easily verify that it approaches $2^{-1}g'^{-}(t_0)$ as y approaches $0+$. Upon using the integration by parts formula, the second term reduces to

$$- sy\int_0^{t_0} h(t)(t - t_0)[R(t) + tR'(t)]dt.$$

If J denotes the value of the last expression, then

$$|J| < \frac{2y}{\pi}\int_0^{t_0} |h(t)|R(t)dt.$$

For each $\epsilon > 0$, there exists $\gamma > 0$ such that $|h(t)| < \frac{1}{2}\epsilon t_0^2$ for $t_0 - t < \gamma$, so that

$$\frac{2y}{\pi}\int_{t_0-\gamma}^{t_0} |h(t)|R(t)dt < \frac{2\epsilon y t_0^2}{\pi}\int_0^1 R(t)dt < \epsilon$$

for $y > 0$. Since $h(t)$ is a bounded function, there exists $\gamma' > 0$ such that, for $y < \gamma'$,

$$\frac{2y}{\pi}\int_0^{t_0-\gamma} |h(t)|R(t)dt < \epsilon.$$

In order to treat the second integral appearing in (2) we replace $g(t)$ by

$$[g(t_0) + g'^{+}(t_0)(t - t_0) + k(t)(t - t_0)]$$

and proceed as above. However, in this case the integration by parts formula yields the additional term

$$\pi^{-1}y[k(1)(1 - t_0)/t_0^2(1 + y^2)]$$

which approaches zero with y. This completes the proof of the Theorem.

Remark added in the revision. I am indebted to the referee for suggesting the revised form of the Theorem. Also, as he points out, the Theorem is valid if we replace the interval of integration $[0, 1]$ by the ray $[0, \infty]$ and restrict z so that $z \notin [-\infty, 0]$.

REFERENCE

1. D. V. Widder, *The Laplace Transform*, (Princeton, 1946).

Illinois Institute of Technology, Chicago

es,
to
o-
cal
er,
nd
he
he
n-
si-
of
2.

he
of
nd

ed
at
ts
95